Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: The Severe Covid-19 GWAS Group. Genomewide association study of severe Covid-19 with respiratory failure. N Engl J Med. DOI: 10.1056/NEJMoa2020283

Table of contents

Supplementary Note

Members of the Humanitas COVID-19 Task Force.

Humanitas Clinical and Research Center – IRCCS –, via Manzoni 56, 20089 Rozzano (Milan) – Italy

Aghemo Alessio, Angelini Claudio, Badalamenti Salvatore, Balzarini Luca, Bocciolone Monica, Cecconi Maurizio, Ciccarelli Michele, Kurihara Hayato, Lagioia Michele, Omodei Paolo, Preatoni Paoletta, Voza Antonio, Accornero Stefano, Alfarone Ludovico, Ali Hussam, Arcari Ivan, Arosio Paola, Azzolini Elena, Baccarin Alessandra, Baggio Sara, Barbagallo Michela, Barberi Caterina, Barbic Franca, Barbieri Viviana, Barbone Alessandro, Basciu Alessio, Benvenuti Chiara, Bianchi Ilaria, Bombace Sara, Borea Federica, Borroni Mario, Bresciani Gianluigi, Brunetta Enrico, Bulletti Cinzia, Cadonati Cristina, Calabro' Loenzo, Calatroni Marta, Calvetta Albania Antonietta, Cannata Francesco, Canziani Lorenzo, Capogreco Antonio, Capretti Giovanni Luigi, Carlani Elisa, Carrone Flaminia, Casana Maddalena, Ceribelli Angela, Ceriotti Carlo, Cimino Matteo, Ciuffini Leonardo, Colaizzi Chiara, Colapietro Francesca, Costa Guido, Cozzi Ottavia, Craviotto Vincenzo, Crespi Chiara, Crippa Massimo, Da Rio Leonardo, Dal Farra Sara, D'antonio Federica, De Ambroggi Guido, De Donato Massimo, De Lucia Francesca, De Nittis Pasquale, De Santis Maria, Delle Rose Giacomo, Di Pilla Marina, Dipaola Franca, Dipasquale Andrea, Dipasquale Angelo, Droandi Ginevra, Fazio Roberta, Ferrante Giuseppe, Ferrara Elisa Chiara, Ferrari Matteo Carlo, Ferri Sebastian, Folci Marco, Foresti Sara, Franchi Eloisa, Fraolini Elia, Furfaro Federica, Galimberti Paola, Galtieri Alessia, Gavazzi Francesca, Generali Elena, Goletti Benedetta, Guidelli Giacomo, Jacobs Flavia, Lania Andrea Gherardo, Libre' Luca, Lleo Ana, Loiacono Ferdinando, Lughezzani Giovanni, Maccallini Marta, Magnoni Paola, Maiorino Alfonso Francesco, Malesci Alberto, Mantovani Riccardo, Marchettini Davide, Marinello Arianna, Markopoulos Nikolaos, Masetti Chiara, Mazziotti Gherardo, Milani Angelo, Mirani Marco, Morelli Paola, Motta Francesca, Mundula Valeria, Nigro Mattia, Ormas Monica, Pagliaro Arianna, Paliotti Roberta, Pavesi Alessia, Pedale Rosa, Pegoraro Francesco, Pellegatta Gaia, Pellegrino Marta, Petriello Gennaro, Piccini Sara, Pocaterra Daria, Poliani Laura, Procopio Fabio, Puggioni Francesca, Pugliese Luca, Pugliese Nicola, Racca Francesca, Randazzo Michele, Regazzoli Lancini Damiano, Reggiani Francesco, Rimoldi Monica, Rodolfi Stefano, Ruongo Lidia, Sacco Clara, Sagasta Michele, Sandri Maria Teresa, Savi Marzia, Scarfo' Iside, Shiffer Dana, Sicoli Federico, Solano Simone, Solitano Virginia, Spata Gianmarco, Stainer Anna, Stella Matteo Carlo, Strangio Giuseppe, Taormina Antonio, Testoni Lucia, Tordato Federica, Trabucco Angela, Ulian Luisa, Valentino Rossella, Valeriano Chiara, Vena Walter, Verlingieri Simona, Vespa Edoardo, Zanon

Veronica, Zanuso Valentina, Zilli Alessandra, Lutman Fabio, Lanza Ezio, Profili Manuel, Giannitto Caterina, Mrakic Sposta Federica, Torrisi Chiara, Poretti Dario, Pedicini Vittorio, Bonifacio Cristiana, D'orazio Federico, D'antuono Felice, Castelli Alice, Pestalozza Alessandra, Paiardi Silvia, Teofilo Francesca Ilaria, Citterio Gianluigi, Aloise Monia, Ripoll Pons Marta, Lavezzi Elisabetta, Fedeli Carlo, Desai Antonio, Caltagirone Giuseppe, Voza Giuseppe, Giorgino Massimo Giovanni

Members of the COVID-19 Host Genetics Initiative.

The COVID-19 Host Genetics Initiative is described in further details at https://www.covid19hg.org/ where an updated version of all members can be found at all times. Named members listed below are acknowledged for their involvement in contributing to the publicly available COVID-19 Host Genetics Initiative summary statistics round 2 (analysis named "20200508-results-ANA5_ALL_inv_var_meta"; file named "COVID19_HGI_ANA5_20200513.txt.gz"; meta-analysis release date of May 15 2020) to which parts of the discussion of the chromosome 3 gene cluster findings of the present study are done (see main text Discussion and **Figure S11**).

Contributor	Study	Affiliation
Lude Franke	Lifelines	Department of Genetics, University Medical Centre Groningen, University of Groningen
Marike H Boezen	Lifelines	Department of Epidemiology, University Medical Centre Groningen, University of Groningen
Patrick Deelen	Lifelines	Department of Genetics, University Medical Centre Groningen, University of Groningen
Annique Claringbould	Lifelines	Department of Genetics, University Medical Centre Groningen, University of Groningen
Esteban A Lopera-Maya	Lifelines	Department of Genetics, University Medical Centre Groningen, University of Groningen
David A van Heel	Genes & Health	Queen Mary University of London, Blizard Institute, 4 Newark Street, London, UK
Bhavi Trivedi	Genes & Health	Queen Mary University of London, Blizard Institute, 4 Newark Street, London, UK
Karen A Hunt	Genes & Health	Queen Mary University of London, Blizard Institute, 4 Newark Street, London, UK
Beverley MacLaughlin	Genes & Health	Queen Mary University of London, Blizard Institute, 4 Newark Street, London, UK
Qinqin Huang	Genes & Health	Wellcome Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK
Hilary C Martin	Genes & Health	Wellcome Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK
Richard C Trembath	Genes & Health	School of Basic and Medical Biosciences, Faculty of Life Sciences and

		Medicine, Kings College London, London, UK
Christopher Griffiths	Genes & Health	Queen Mary University of London, Blizard Institute, 4 Newark Street, London, UK
Sarah Finer	Genes & Health	Queen Mary University of London, Blizard Institute, 4 Newark Street, London, UK
John Wright	Genes & Health	Bradford Institute for Health Research, Bradford Teaching Hospitals National Health Service (NHS) Foundation Trust, Bradford, UK
Dan Mason	Genes & Health	Bradford Institute for Health Research, Bradford Teaching Hospitals National Health Service (NHS) Foundation Trust, Bradford, UK
Finngen	Finngen	Institute for Molecular Medicine Finland, University of Helsinki
Arden Moscati	BioMe	The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY
Judy H. Cho	BioMe	The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY
Ruth J.F. Loos	BioMe	The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY
Elizabeth T. Cirulli	Helix Exome+ COVID-19 Phenotypes	Helix
Kelly M. Schiabor Barrett	Helix Exome+ COVID-19 Phenotypes	Helix
Stephen Riffle	Helix Exome+ COVID-19 Phenotypes	Helix
Alexandre Bolze	Helix Exome+ COVID-19 Phenotypes	Helix
Simon White	Helix Exome+ COVID-19 Phenotypes	Helix
Francisco Tanudjaja	Helix Exome+ COVID-19 Phenotypes	Helix
Xueqing Wang	Helix Exome+ COVID-19 Phenotypes	Helix

Jimmy M. Ramirez III	Helix Exome+ COVID-19 Phenotypes	Helix
Yan Wei Lim	Helix Exome+ COVID-19 Phenotypes	Helix
Nicole L. Washington	Helix Exome+ COVID-19 Phenotypes	Helix
James T. Lu	Helix Exome+ COVID-19 Phenotypes	Helix
Eco de Geus	Netherlands Twin Register	Department of Biological Psychology, Vrije Universiteit Amsterdam
Meike Bartels	Netherlands Twin Register	Department of Biological Psychology, Vrije Universiteit Amsterdam
Jouke-Jan Hottenga	Netherlands Twin Register	Department of Biological Psychology, Vrije Universiteit Amsterdam
Daniel J Wilson[*]	UK Biobank	University of Oxford
Nicolas Arning	UK Biobank	University of Oxford
Jacob Armstrong	UK Biobank	University of Oxford
Justine Rudkin	UK Biobank	University of Oxford
Gavin Band	UK Biobank	University of Oxford
Sarah G Earle	UK Biobank	University of Oxford
Shang-Kuan Lin[#]	UK Biobank	University of Oxford
Derrick W Crook	UK Biobank	University of Oxford
David H Wyllie	UK Biobank	Public Health England
Anne Marie O'Connell	UK Biobank	Public Health England
Tomoko Nakanishi	UK Biobank	Department of Human Genetics, McGill University, Montréal, Québec, Canada
Vincenzo Forgetta	UK Biobank	Centre for Clinical Epidemiology, Department of Medicine, Lady Davis Institute for Medical Research, Jewish General Hospital, McGill University, Montréal, Québec, Canada
Guillaume Butler-Laporte	UK Biobank	Department of Human Genetics, McGill University, Montréal, Québec, Canada
J. Brent Richards	UK Biobank	Department of Human Genetics, McGill University, Montréal, Québec, Canada
Konrad J. Karczewski	UK Biobank	Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA
Elizabeth G. Atkinson	UK Biobank	Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA
Masahiro Kanai	UK Biobank	Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA

Nikolas Baya	UK Biobank	Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA
Patrick Turley	UK Biobank	Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA
Raymond K. Walters	UK Biobank	Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA
Duncan S. Palmer	UK Biobank	Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA
Sam Bryant	UK Biobank	Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA
Claire Churchhouse	UK Biobank	Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA
Hilary Finucane	UK Biobank	Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA
Alicia R. Martin	UK Biobank	Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA
Scott Weiss	Mass General Brigham Biobank	Division of Rheumatology, Inflammation and Immunity, Department of Medicine, Brigham and Women's Hospital
Elizabeth W. Karlson	Mass General Brigham Biobank	Brigham and Women’s Hospital
Jordan W. Smoller	Mass General Brigham Biobank	Center for Genomic Medicine, Massachusetts General Hospital
Susan A. Slaugenhaupt	Mass General Brigham Biobank	Center for Genomic Medicine, Massachusetts General Hospital
Robert Green	Mass General Brigham Biobank	Division of Genetics, Department of Medicine, Brigham and Women’s Hospital, Broad Institute of MIT and Harvard, Harvard Medical School
Yen-Chen Anne Feng	Mass General Brigham Biobank	Center for Genomic Medicine, Massachusetts General Hospital
Josep M. Mercader	Mass General Brigham Biobank	Programs in Metabolism and Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA and Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

Shawn N. Murphy	Mass General Brigham Biobank	Department of Neurology, Massachusetts General Hospital
James B. Meigs	Mass General Brigham Biobank	Division of General Internal Medicine, Massachusetts General Hospital and Dpeartment of Medicine, Harvard Medical School and Program in Medical and Population Genetics, Broad Institute
Ann Woolley	Mass General Brigham Biobank	Brigham and Women's Hospital
Emma Perez	Mass General Brigham Biobank	Division of Genetics, Department of Medicine, Brigham and Women's Hospital
Karolina Chwialkowska	POLCOVID-Genomika	Centre for Bioinformatics and Data Analysis, Medical University of Bialystok, Bialystok, Poland
Margherita Francescatto	The FVG COVID-19 Project + Italian Network of Genetic Isolates	Department of Medical, Surgical and Health Sciences, University of Trieste, Trieste, Italy
Amy Trankiem	N/A	Broad Institute
Matthew Solomonson	N/A	Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA
Rachel Liao	N/A	Broad Institute
Huy Nguyen	N/A	Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA
Juha Karjalainen	N/A	Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA
Wei Zhou	N/A	Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA

[*]This research has been conducted using the UK Biobank Resource under Application Number 53100. DJW is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (Grant 101237/Z/13/B). DJW is supported by the Robertson Foundation.

[#]This work was supported by the Wellcome Trust (BST00080).

Supplementary Methods

Authorship contributions.

T.H.K conceived and initiated the project. T.H.K and A.F. jointly designed, provided infrastructure to and jointly supervised the project. T.H.K, A.F., J.R.H., T.F., D.E. and F.D. wrote the first draft of the manuscript. A.F., D.E. and F.D. vouches for the genetic data and coordinated and performed the statistical analyses with contributions from G. H-S., J.K., M.Wi., M.C-R., L.W., F.U-W. and M.We. S.M., X.Y., A.T., A.Pesc., H.E., M.E.F.B., M.Scha., N.B., O.Ö.,T.L.L., T.W., W.A., and W.P. performed sample processing, performed DNA extraction and/or genotyping. L.V., L.B., M.But., A.Alb., P.I., R.A., J.F., J.M.B., M.R.G., M.M.G. and J.R.H. organised and supervised patient inclusion, vouches for the clinical data, and provided input to study design and the manuscript. M.D.Am., S.D., D.Pra., G.B., C.M., F.A., T.Z., A.B.O., A.G.C., A.Pese., A-E.G-F., A.B-G., A.Z., A.Ban., A.Pes, A.A., A.Ll., A.Bio., A.C-G., A.G., A.C.N., A.La., A.L-F., A.J., A.Pal., A.Pro., A.V., A.S., B.M., B.N.J., C.Q., C.P., C.G., C.A., C.C., D.J., D.P., E.M.D., E.S., E.M.P., E.N., F.G.S., F.C., F.M-B., F.P., F.B., F.M., F.R-F., G.G., G.Cos., G.Car., G.F., G.M., H.K., I. G-F., I.M., J.M., J.E., J.F-A., K.G-E., L.I-S., L.R.B., L.Tél., L.M., L.Sa., L.Su., L.Te., L. Sc., L.R., M.R.B, M.G.V., M.H-T., M.A-H., M.D.An., M.Bal., M.Car., M.Caz., M.Cic., M. Cec., M.R-G., M.Boc., M.M., N.Ma., N.Mo., N.S., O.Ö., O.P., P.F., P.P., P.B., P.O., P.T., P.C., P.M.R., R.D.C., R.D.P., R.F., R.G, R.N., S.Bad., S.G., S.M., S.J., S.A., S.P., S.Bos., The Humanitas COVID-19 Task Force, T.P., T.L.L., V.R., V.Mon. and V.Mor. provided samples, phenotypic data and intellectual input to the manuscript. M.Schu. contributed to the manuscript. The decision to publish was made by D.E., F.D., L.V., L.B., M.But., A.Alb., P.I., R.A., J.F., J.M.B., M.R.G., J.R.H., T.F., A.F. and T.H.K in agreement with all authors. All authors revised and edited the manuscript for critical content and approved of the final version to be published.

DNA extraction and genome-wide SNP genotyping.

DNA extraction was performed by the DNA laboratory of the Institute of Clinical Molecular Biology (Christian-Albrechts-University of Kiel, Germany) using a Chemagic 360 from PerkinElmer (Waltham, Massachusetts, U.S.) with the low volume kit cmg 1491 and the buffy coat kit cmg-714 (Chemagen, Baesweiler, Germany) according to the manufacturer's protocol. The Chemagen chemistry for DNA extraction from whole blood and buffy coat is based on the use of magnetic beads. Up to 400 µl whole blood or 300 µl buffy coat were

used for isolation, depending on the shipped material of the different centers. In a first step, the cell lysis for protein degradation by protease is performed. The isolation of the DNA is achieved by capturing the DNA with polyvinyl alcohol magnetic beads (M-PVA Magnetic Beads). The DNA is bound to the surface coating of these beads. These beads, together with the bound DNA, are attracted by the magnetized metal rods, which can then transfer the DNA from one wash buffer to another. After deactivation of the electromagnet, the particles are resuspended in the solution. Finally, beads were transferred into 100-250 µl elution buffer, which inactivates the interaction between the beads and the DNA. The magnetic beads are removed, leaving the isolated DNA in suspension. The concentration was measured in a Dropsense96 (Unchained Labs, Pleasanton, CA, U.S.) spectrophotometer.

Genotyping of all patient panels and the control panel genotyped for the purpose was conducted by the Institute of Clinical Molecular Biology's DNA Laboratory and Genotyping Core Facilities, employing Illumina's (Illumina Inc., San Diego, U.S.) Global Screening Array-24 Multi Disease (GSA) Version 2.0 B1 following the Illumina(R) Infinium HTS Assay Auto 3-day Workflow (Document #15045738v0). In brief, initial DNA quantification, amplification and incubation for 24 hours was performed according to protocol on day 1. Next, we performed enzymatic DNA fragmentation, followed by 2-propanol precipitation, resuspension and overnight hybridization of DNA to the BeadChip on day 2. Last, BeadChip washing removing unhybridized and nonspecifically hybridized DNA, extension adding labeled nucleotides to extend primers hybridized to the sample, staining of primers and final imaging using the Infinitum LCG scan setting was performed following the manufacturer's protocol on day 3.

The genome-wide content of the GSA was selected by the vendor for high imputation accuracy at minor allele frequencies of >1% across all twenty-six 1,000 Genomes Project populations.[1] The clinical research content includes 712,189 variants with established disease associations, relevant pharmacogenomics markers, and curated exonic content based on ClinVar, NHGRI, PharmGKB, and ExAC databases.

External Spanish and Italian GWAS control data sets.

For later validation of allele frequencies in the independent Spanish and Italian GWAS control data sets, we collected six additional external control data sets from different non-Illumina GSA genotyping platforms:

External Italian controls 1: n=952 genotyped controls (data from the Illumina Hap550k array) from references [2] and [3]. External Italian controls 2: n=3,361 genotyped controls (data from the Affymetrix 6.0 array) from reference [4]. External Spanish controls 1: n=1,482 genotyped controls (data from the Human610-Quad array) from reference[5]. External Spanish controls 2 and 3: n=1,617 genotyped controls (data from the Illumina HumanCNV370k and Illumina HumanCore arrays, respectively) from reference [6]. External Spanish controls 4: n=4,988 genotyped controls (data from the Illumina Infinium Multi-Ethnic Global-8 v1.0 array) from references [7] and [8].

In **Table S9** the German blood donors of the UKSH Cohort were persons donating at least once a unit of whole blood between 2006 and 2019 at University Hospital Schleswig-Holstein (UKSH). Of 104,604 donors, 62,267 were first time donors in that time frame. 110,309 patients receiving at least one unit of red blood cells (RBC) between 2006 and 2019 were also included. The Italian and Spanish blood donors were persons donating blood at the blood transfusion service (BTS) of the IRCCS Fondazione Cà Granda Ospedale Maggiore, Milan, Italy (first time blood donors from 2015 to 2019, total n = 14,658), the BTS Centro de Transfusion de Madrid, Spain (first time blood donors in 2019, total n = 36,781) and from the BTS Banc de Sang i Teixits, Barcelona, Spain, (first time blood donors in 2019, total n = 34,774).

Genotype calling, SNP and sample quality control and principal component analysis.
Initial genotype calling extracting GSA genotyped data from intensity data files was performed with the Illumina GenomeStudio v. 2.0 software with the cluster definition file GSAMD24v2-0_20024620_A1-762Samples-LifeBrain (https://www.illumina.com/documents/products/datasheets/datasheet_genomestudio_software.pdf. Finally, we had 712,189 SNPs before quality control (QC). Based on initial genotype data, we removed samples with <90% callrate using PLINK.[9]
After genotype calling a unified QC procedure was carried out for the Spanish and Italian case-controls GWAS data sets. Variants that had >2% missing data, a minor allele frequency (MAF) <0.1% in either of the different disease sets or in controls, different missing genotype rates in affected and unaffected individuals ($P_{Fisher}<10^{-5}$) or deviated from Hardy-Weinberg equilibrium (with a false discovery rate (FDR) threshold of 10^{-5} in controls) (a) across the entire collection with at most one batch being removed or (b) falling below in two single batches, were excluded. Samples that had >2% missing data and overall increased/decreased

heterozygosity rates (i.e. ±5 SD away from the sample mean) were removed. For robust duplicate/relatedness testing (IBS/IBD estimation) and population structure analysis, we used a linkage disequilibrium (LD)-pruned subset of SNPs on the basis of a set of independent (MAF>5%) SNPs excluding X- and Y-chromosomes, SNPs in LD (leaving no pairs with r^2>0.2), and 11 high-LD regions as described by Price *et al.*[10] Pairwise percentage IBD values were computed using PLINK. By definition, Z0: P(IBD=0), Z1: P(IBD=1), Z2: P(IBD=2), Z0+Z1+Z2=1, and PI_HAT: P(IBD=2) + 0.5 * P(IBD=1) (proportion IBD). One individual (the one showing greater missingness) from each pair with PI_HAT>0.1875 was removed. A value of 0.1875 (proportion IBD) corresponds to a theoretical relationship of halfway between the expected IBD for third- and second-degree relatives.

To identify ancestry outliers, i.e. subjects of non-European ancestry, and to resolve within-Europe relationships and to test for population stratification within and across batches (merged for the Italian and Spanish panels), we performed principal component analysis (PCA) for remaining QCed cases and controls using the PCA method, as implemented in FlashPCA[11], using a LD-pruned subset of SNPs (see text above). Ancestry outliers not matching European populations were removed (**Figure S1, A** and **B**). After QC, PCA revealed no non-European ancestry outliers (**Figure S1, C** and **D**) when performing PCA including reference samples from the 1,000 Genomes reference panel.[1]

We applied same QC procedure as described above to the additional (previously elsewhere genotyped) Spanish and Italian GWAS control data sets from different Illumina and Affymetrix genotyping platforms, leaving 8,036 Spanish and 4,273 Italian controls after QC (**Table S3**).

Genotype imputation.

The QCed Italian and Spanish GWAS datasets comprised 835 Italian Covid-19 cases, 1,255 Italian controls, 775 Spanish Covid-19 cases and 950 Spanish controls (**Table S1B**), and contained 664,571 (Italy) and 662,223 (Spain) variants after QC and filtering of SNPs with alleles AT or CG (the latter often leading to strand issues during imputation). Genotype imputation was conducted for chromosomes 1-22 and X data using the novel TOPMed Freeze5 on genome build GRCh38 and the Michigan Imputation Server (https://imputation.biodatacatalyst.nhlbi.nih.gov/index.html#!).[12] We provided the input data in “vcf.gz” format as GRCh38 build. We used the offered population panel “ALL” and

applied the server-side option to filter by an imputation R^2 with threshold 0.001. The final imputed results contained 186,310,922 variants in the Italian dataset and 157,331,047 variants in the Spanish dataset after TOPMed imputation. For the imputation of the X chromosome, we coded males as diploid in the non-pseudoautosomal (non-PAR) region.[13] After quality control and imputation using TOPMed in total 8,965,091 SNPs were included for the Italian panel and 9,140,716 SNPs for the Spanish panel, with post imputation $R^2 \geq 0.6$ and MAF ≥1%.

The genotype imputation procedure described above was applied also to the independent and QCed Spanish and Italian GWAS control data sets from different Illumina and Affymetrix genotyping platforms (**Table S3**).

Next generation sequencing-based HLA allele typing.

HLA genotypes were typed using NGS-sequencing at full-context 3-field (5'UTR-3'UTR) resolution for *HLA-A, -B, -C, -DQA1, -DQB1, -DPB1*. Sequencing for *HLA-DRB1* was conducted only for exon 2 – 3'UTR. Sequencing was performed with a MiSeq (paired-end with a length of 250 bp and 500 sequencing cycles). For *HLA-DPB1* G groups were assigned for alleles where phasing of heterozygote positions in exon2 and exon3 could not be resolved. The required target enrichment step prior to the library preparation was accomplished by locus specific long-range polymerase chain reactions (PCRs). The primer sets, which were designed and validated in-house, target (with the exception of the *DRB1* forward primer) sequence motifs within the UTRs of the HLA loci. In addition, the size of the *HLA-DPB1* gene required a primer design with two overlapping amplicons to cover the entire gene. A unified PCR setup based on the GoTaq® Long PCR Master Mix (Promega) and the recommendations of the manufacturer was applied to amplify the HLA genes. To quality control the amplification step, all HLA amplicons were monitored on 1% TAE agarose gels. After electrophoresis, all HLA amplicons of the individual samples were equimolarly pooled equally into their corresponding wells. Subsequent steps of the entire library preparation from enzymatic fragmentation, end repair, adapter ligation to paired end Index PCR were done with the NEBNext®Ultra™DNA Library Prep Kit for Illumina (NEB) according to the manufacturer's instructions. Prior to each MiSeq (Illumina) run, we pooled up to 192 individual amplicon fragment pools, after having size selected library fragments between 600-800 bp with an electrophoresis-based method (1.6% TAE agarose gel separation, GeneJet Gel Extraction Kit, ThermoFisher). For quantification, we used the Kapa Library

Quantification Kit (Roche) according to the manufacturer supplied protocol. Before loading to the MiSeq cartridges the sequencing libraries were diluted (15pM) and denatured according to standard Illumina MiSeq loading procedures.

Paired-end next generation sequencing (NGS) of the size-selected HLA amplicon library was performed on a MiSeq (Illumina) running 500 cycles of V2 chemistry. The generated fastq data files were analyzed in parallel with the HLA Twin Software version 4.2.1 (Omixon) and the NGSengine version 2.16.2 (GenDx).

The IPD-IMGT/HLA database release version 3.38.0 was used as the reference for alignment to allow accurate comparison of alleles genotyped using both HLA software tools. A parameter for acceptance of HLA genotyping results was an average depth of coverage ≥ 100x.

For the HLA analysis, the 3-digit typings were translated to 2-digit G groups using the hla_nom_g.txt from (http://hla.alleles.org/alleles/g_groups.html; downloaded 05/2020).

Genome-wide association analysis.

Covid-19 patients with respiratory failure (cases) were compared with population controls (unknown Covid-19 status). Case-control allele-dose association tests of the genotyped and imputed SNPs in the Italian and Spanish panel were performed separately using PLINK 1.9 (logistic regression).[9] Two different analyses including covariates from the PCA without 1,000 Genomes reference samples were run to control for potential population stratification (analysis I) as well as potential population stratification and potential age and gender effects (analysis II):

Analysis (I) case/control status ~ SNP + PC1 + PC2 + PC3 +PC4 + PC5 + PC6 + PC7 + PC8 + PC9 + PC10
Analysis (II) case/control status ~ SNP + age + gender + PC1 + PC2 + PC3 +PC4 + PC5 + PC6 + PC7 + PC8 + PC9 + PC10

Only high-quality variants with an imputation score $R^2 \geq 0.6$ and MAF≥1% were considered. Overall, 9,140,716 variants were analysed for Spain and 8,965,091 variants were analysed for Italy.

Meta-analysis of Italian and Spanish GWAS datasets.

A fixed-effects meta-analysis was conducted using the meta-analysis tool METAL[14] (https://genome.sph.umich.edu/wiki/METAL_Documentation) on 8,582,968 variants overlapping between both studies with $R^2 \geq 0.6$, using the BETA and its standard error (SE) from the study specific association. Using METAL, the combined P-value and the combined effect (E) with its SE was calculated from which we computed the odds ratio (OR) and its 95% confidence interval (CI; OR = exp(E), the upper confidence limit (OR_95U) was calculated as OR_95U=exp(E+1.96*SE), the lower confidence limit (OR_95L) was calculated as OR_95L=exp(E-1.96*SE)). Quantile-quantile plot of the cross-country meta-analyses and country-wise association analyses showed only excess of significant associations in the tail of the distribution (**Figures S2** and **S3**).

Identification of genomic loci showing suggestive association.

We used FUMA[15] to identify genomic loci showing suggestive evidence ($P<1x10^{-5}$) for association with severe Covid-19 with respiratory failure (**Table S5**). According to FUMA, "independent significant" SNPs are defined as SNPs that have a P-value < $1x10^{-5}$ and are independent from each other at the linkage disequilibrium threshold $r^2<0.6$. Therefore, independent significant SNPs are essentially the same as SNPs that are contained after clumping GWAS tagged SNPs at the same P-value and r^2. Independent significant SNPs are used to select candidate SNPs that are in LD with the independent significant SNPs. "Lead SNPs" are defined as SNPs which are independent significant SNPs and are independent from each other at LD $r^2 < 0.1$. In addition to lead SNPs, FUMA defines "genomic risk loci", including all independent signals that are physically close or overlapping in a single locus. First, independent significant SNPs that are dependent each other at $r^2 \geq 0.1$ are assigned to the same genomic risk locus. Then, independent significant SNPs that are closer than 250 kb are merged into one genomic risk locus. The distance between two LD blocks of two independent significant SNPs is the distance between the closest SNPs (which are in LD of the independent significant SNPs at user defined r^2) from each LD block. Each locus is represented by the top lead SNP that has the minimum P-value at the locus.

Bayesian fine-mapping analysis.

A Bayesian fine-mapping analysis was carried out using FINEMAP[16] in order to determine a credible set, i.e. a minimum set of variants containing all causal variants with certainty ≥0.95%, and to calculate the posterior inclusion probability (PIP) for each SNP as causal in

any of the models. To this end, we extracted regions of interest from the imputed genotype data and calculated the local LD structure in the Italian and Spanish discovery sets, which served together with the meta-analysis I summary statistics as an input for FINEMAP. We set the option --n-causal-snps 1.

ABO blood group SNP genotyping and analysis.

By using SNP genotypes extracted from the Spanish and Italian TopMED genotype imputation, we performed a SNP based ABO typing to identify the core ABO blood groups A, B, AB or O.

We extracted the genotypes of three SNPs: rs8176719, rs41302905 and the adjacent rs8176747, further described below an inferred blood group status based on these SNPs. For blood group allele O, two different types exist, that are named O1 (more common) and O2 (rare). The genotyped variants for rs8176719 were used to infer blood group O1 status. rs8176719 is a deletion, that encodes for the common O-allele (O1) and is the causative SNP for chain termination at exon 6 (261delG causes Thr88Profs*31), which determines that an individual has this variation of the ABO transcript (allele). Genotyped variants at rs41302905 were used to infer for allele O2 status. Here, the rare A nucleotide encodes for the 2nd frequent O-allele (O2) and is the causative SNP (802G>A causes Gly268Arg), which determines that an individual carries this allele. Genotypes for rs8176747 were used to infer allele B status. Here, the rare C nucleotide encodes for allele B and is the causative SNP (803G>C causes Gly268Ala), which determines if an individual has this B-allele.[17] Individuals with the diploid alleles O1O1, O1O2 and O2O2 were assigned blood group O, individuals with diploid alleles BO1, BO2 and BB were assigned blood group B. We type blood group allele A by exclusion, assuming that an individual has allele A, if neither B or O were assigned by the criteria above. Individuals with a diploid alleles AO1, AO2 and AA were assigned blood group A, and individuals with diploid allele AB were assigned blood group AB.

We calculated A, B, AB, O blood group frequencies in cases and controls, respectively. A logistic regression was conducted on these data in two adjusted analyses (analysis I and II) to control for (I) potential population stratification as well as (II) potential population stratification as well as age and gender effects. We additionally performed unadjusted and adjusted analyses on the cases only, with the independent variable binarized to no

mechanical oxygenation vs. mechanical oxygenation. Blood group specific odds ratios were calculated by making four different grouping as A vs. B/AB/O, B vs. A/AB/O, AB vs. A/B/O and O vs. A/AB/B. To combine study specific estimates, a fixed-effects meta-analysis was performed using the R-package metafor on the beta estimates and corresponding standard errors. Association test statistics were weighted by the number of samples in each sub-analysis per cohort.

Tissue specific eQTL data of lead SNPs of genome-wide significant loci and their proxies.

Tissue specific expression quantitative trait loci (QTLs) of GWAS lead rs657152 and rs11385942 variants as well as their respective proxies, rs545971 and rs17214952, were obtained from GTEx[18] and conditional eQTL catalogue[19] databases. The API service was used to retrieve eQTLs of rs11385942 and rs17214952 variants from the GTEx v8. Since the later version of GTEx does not contain information on rs657152 and rs545971, their eQTL data was retrieved from GTEx v7. Additional data on rs17214952 variant was obtained from conditional cis-eQTL mapping catalog for whole blood.

Gene expression data of genome-wide significant loci overlapping candidate genes.

Bulk mRNA consensus tissue gene expression data and bulk blood cell gene expression data were obtained from the Human Protein Atlas (HPA) database.[20] This database includes data from three sources: HPA, GTEx[18] and FANTOM5.[21] The obtained normalized gene expression values were gene-wise centered and z-score normalized for visualization. Single cell RNA-Sequencing (sc-RNA-Seq) data for SARS-CoV-2 relevant tissues were obtained from the COVID-19 Cell Atlas.[22] Lung and upper airway sc-RNA-Seq datasets[23,24] were retrieved in .h5ad format files and visualized using the scanpy v1.4.6 package.[25]

HLA next generation sequencing data association analysis.

The associations at the HLA complex at 6p21 were refined by performing sequencing-based HLA typing of 7 HLA loci (*HLA-A, -C, -B, -DRB1, -DQA1, -DQB1, -DPB1*) in a subset of 835 cases and 891 controls from Italy and 773 cases from Spain. Allele association analysis was performed on case-controls for the Italian samples and on disease severity in cases only. Severity was binarized to no mechanical oxygenation vs. mechanical oxygenation. Logistic regression was conducted in two adjusted analyses (analysis I and II) to control for (I) potential population stratification as well as (II) potential population stratification as well as

age and gender effects. HLA alleles were coded as present (P) or absent (A) and coded as 0 for AA, 1 for AP and 2 for PP.

We also calculated allele numbers across the classical class I (*HLA-A, -B, -C*) and class II (*HLA-DRB1, -DQB1*) loci as a measure of multilocus heterozygosity (ranging from 3 and 2 alleles for complete homozygosity to 6 and 4 for complete heterozygosity, respectively). A more nuanced measure of allelic variability, the amino acid sequence divergence among alleles, was also calculated for each of the five HLA loci separately and across all three class I or two class II loci, using the GranthamDist tool[26] This measure, also called HLA evolutionary divergence (HED) has proven to predict control of HIV replication[27] as well as response to cancer immunotherapy.[28] Association between these compound parameters of genetic variability and disease risk (cases vs. control) was tested using the same logistic regression model as for the SNP and HLA genotypes (analysis II), including gender, age and the first 10 PCs as covariates. For analysis of the combined data, we also included country (Spain/Italy) as a covariate to account for differences between the datasets.

HLA peptide binding prediction.

Due to the substantial overlap in bound peptides among HLA alleles and their co-dominant expression, we considered the possibility that the additive GWAS association test would not capture the entire functional role of HLA in Covid-19 risk. We therefore employed computational in-silico binding prediction of the observed HLA alleles for SARS-CoV-2 peptides, based on the NGS-based targeted genotyping (see above). Specifically, we aimed to calculate the number of SARS-CoV-2 peptides an individual's HLA alleles are predicted to bind. In order to calculate the number of bound peptides by each allele we used the HLA-peptide binding prediction algorithms netMHCpan (v4.1) and netMHCIIpan (v4.0) for HLA class I and class II alleles, respectively.[29] Analysis of proteome sequences from the GISAID database[30] revealed that the strain G was responsible for more than half of the infections in both Italy and Spain during February and March 2020. Hence, we used a proteome belonging to strain G (EPI_ISL_4174479)[31] to infer all possible potentially relevant peptides (9mers for class I and 15mers for class II). Default %Rank_EL thresholds were used to define strong (0.5% for netMHCpan and 2% for netMHCIIpan) and both weak and strong (2% for netMHCpan and 10% for netMHCIIpan) binders. The total number of bound peptides per individual was calculated for each locus as well as for all class I and class II variants together, respectively, given the individual's HLA genotypes. According to Grifoni

et al.[32], CD4+ T-cell responses concentrate on M,N and Spike proteins and CD8+ T-cell responses on M,NSP6 and Spike proteins. Therefore, we also calculated specifically the number of bound peptides by class I alleles from M, NSP6 and Spike proteins and by class II alleles from M, N and Spike. Finally, the calculations were repeated by using only the Spike protein, given its importance in the transmission of SARS-CoV-2. Association between predicted number of bound peptides and disease risk (cases vs. control) was tested using the same logistic regression model as for the SNP and HLA genotypes (analysis II), including gender, age and the first 10 PCs as covariates. For analysis of the combined data, we also included country (Spain/Italy) as a covariate to account for differences between the datasets.

Supplementary Results

HLA analyses.

We found neither significant differences in allelic distribution between patients and controls in Italy nor between patiens with oxygen supplementation only and those with mechanical ventilation of any kind (**Table S10**). Neither the number of classical HLA alleles of an individual nor the HLA allele divergence at specific loci or across multiple loci differed significantly between cases and controls or between groups of cases with different respiratory support (GLMs, all P-values >0.1). When computationally predicting individual HLA-binding of SARS-CoV-2 peptides, we found no robust statistically significant association between the different peptide values and either disease risk (case vs. control) or disease severity (different levels of respiratory support). Only the predicted total number of peptides from the M, NSP6 and Spike proteins bound by an individual's HLA class I alleles showed a nominally significant variation among the four levels of respiratory support when tested across the entire dataset (Kruskal-Wallis test, P=0.042, not corrected for multiple-testing), driven by a slightly lower number of bound peptides in the most severe group, compared to the other groups (GLM, X=4.3, P=0.038). However, the number of cases in this most severe treatment group (n=6) was too small to draw robust conclusions from this result.

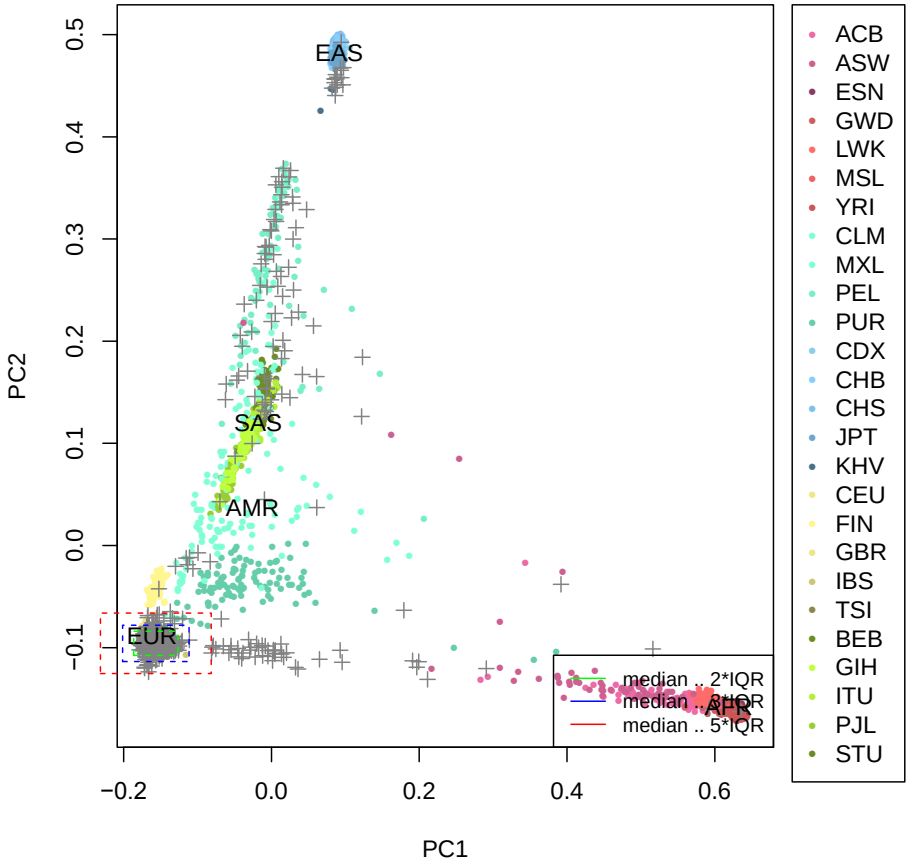
Supplementary Figures

Figure S1. Principal components analyses of Covid-19 cases and controls.
Scatter plots of the principal component analysis (PCA) for cases and controls using the PCA method as implemented in FlashPCA[11], using an LD-pruned subset of SNPs (see **Supplementary Methods**). Ancestry outliers not matching European populations were removed (**A and B**). After QC, PCA revealed no non-European ancestry outliers (**C and D**) when performing PCA including reference samples from the 1,000 Genomes reference panel.[1]

(A) Italian cases and controls before exclusion of ancestry outliers.
(B) Spanish cases and controls before exclusion of ancestry outliers.
(C) Italian cases and controls after exclusion of ancestry outliers.
(D) Spanish cases and controls after exclusion of ancestry outliers.

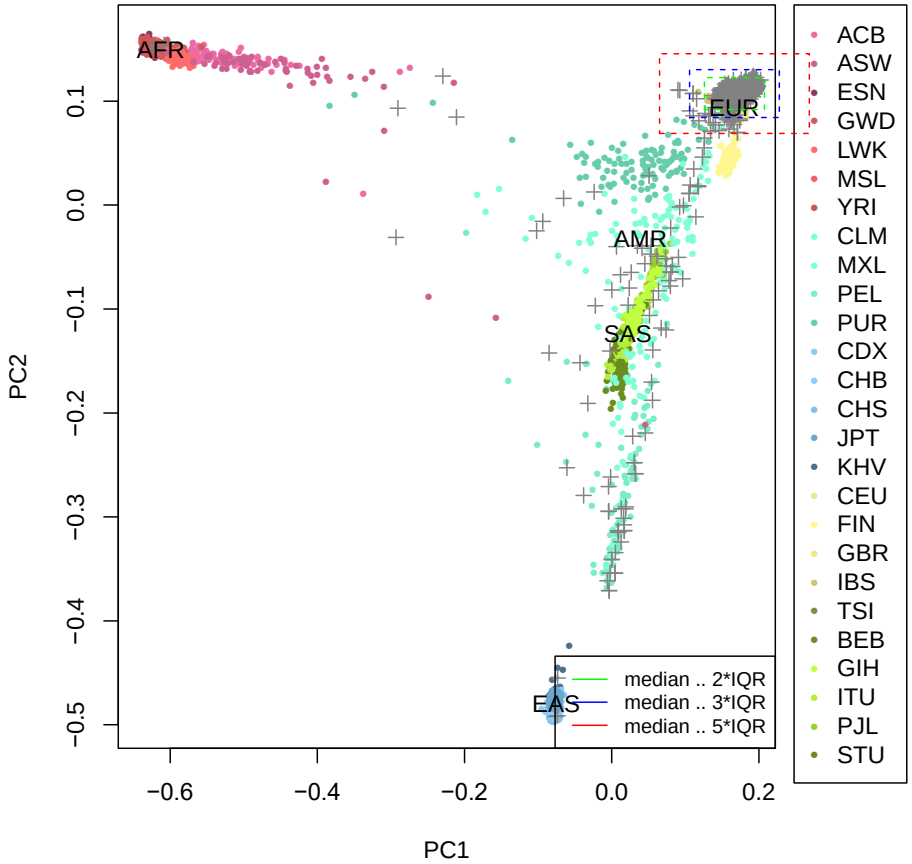
The grey crosses represent Covid-19 cases and controls. The coloured points represent the five super populations retrieved form the 1,000 Genomes data:[1]
African (AFR, purple), Ad Mixed American (AMR, turquoise), South Asian (SAS, green), East Asian (EAS, blue), (EUR, yellow). Population code (super population code): CHB (EAS) Han Chinese in Beijing, China; JPT (EAS) Japanese in Tokyo, Japan; CHS (EAS) Southern Han Chinese; CDX (EAS) Chinese Dai in Xishuangbanna, China; KHV (EAS) Kinh in Ho Chi Minh City, Vietnam; CEU (EUR) Utah Residents (CEPH) with Northern and Western European Ancestry; TSI (EUR) Toscani in Italia; FIN (EUR) Finnish in Finland; GBR (EUR) British in England and Scotland; IBS (EUR) Iberian Population in Spain; YRI (AFR) Yoruba in Ibadan, Nigeria; LWK (AFR) Luhya in Webuye, Kenya; GWD (AFR) Gambian in Western Divisions in the Gambia; MSL (AFR) Mende in Sierra Leone; ESN (AFR) Esan in Nigeria; ASW (AFR) Americans of African Ancestry in SW USA; ACB (AFR) African Caribbeans in Barbados; MXL (AMR) Mexican Ancestry from Los Angeles USA; PUR (AMR) Puerto Ricans from Puerto Rico; CLM (AMR) Colombians from Medellin, Colombia; PEL (AMR) Peruvians from Lima, Peru; GIH (SAS) Gujarati Indian from Houston, Texas; PJL (SAS) Punjabi from Lahore, Pakistan; BEB (SAS) Bengali from Bangladesh; STU (SAS) Sri Lankan Tamil from the UK; ITU (SAS) Indian Telugu from the UK.

(A) Italian cases and controls before exclusion of ancestry outliers.



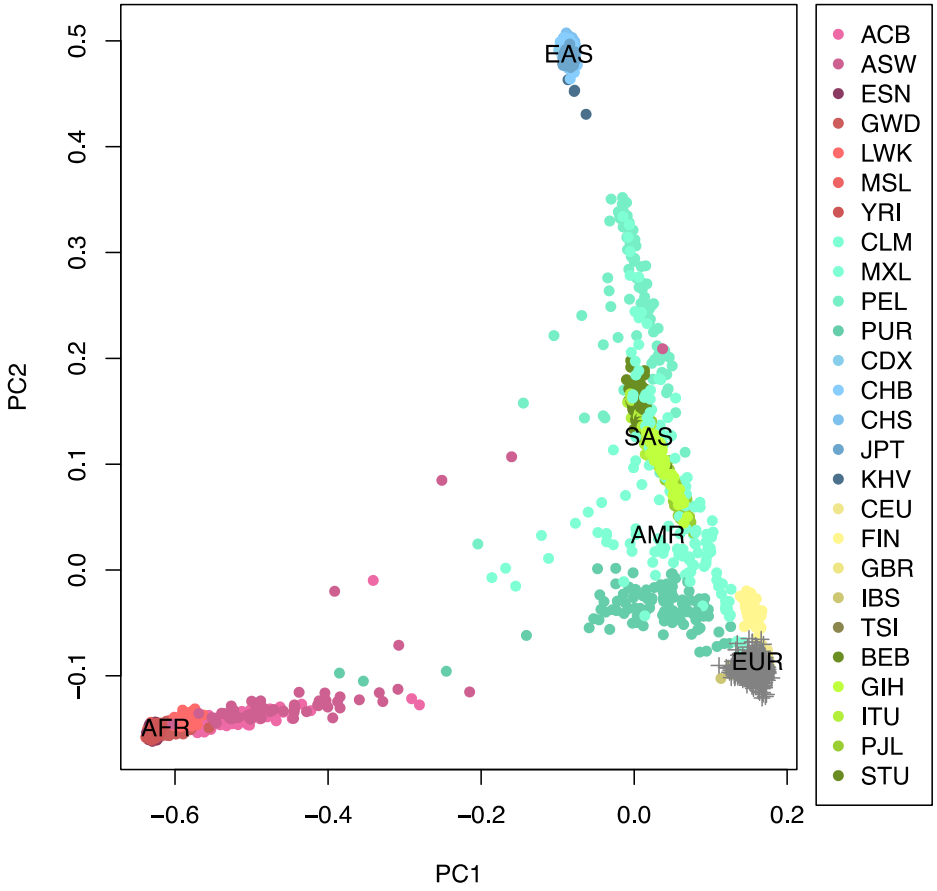
PC1: principal component 1, PC2: principal component 2

(B) Spanish cases and controls before exclusion of ancestry outliers.



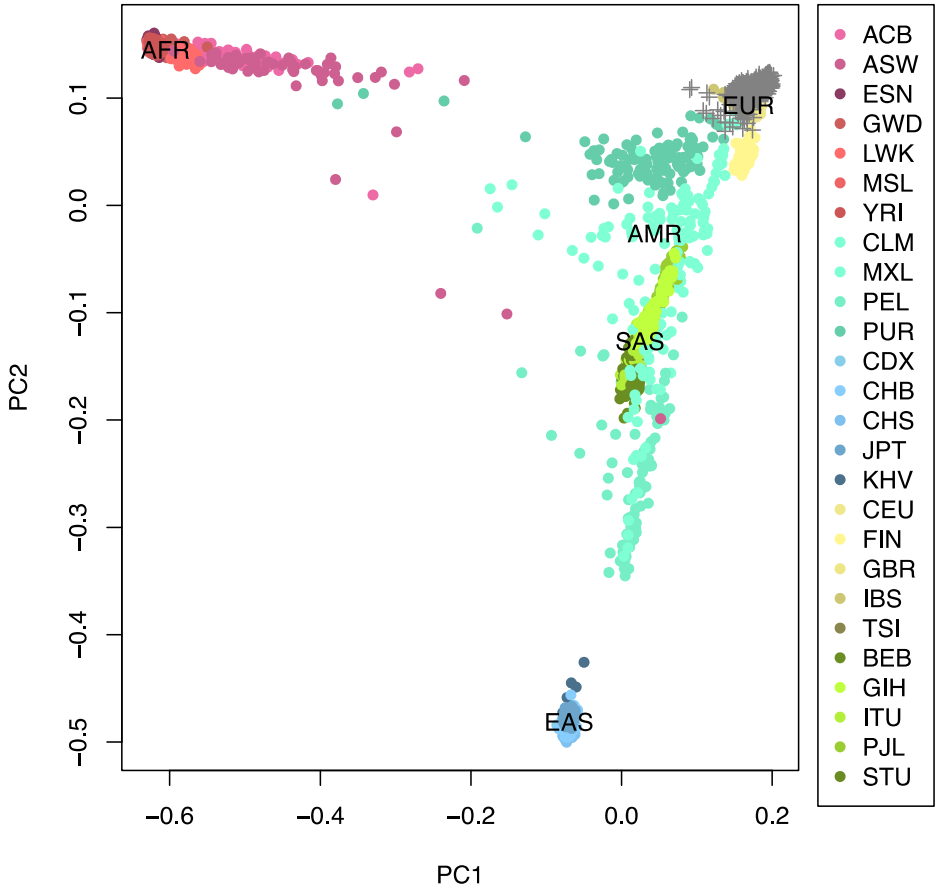
PC1: principal component 1, PC2: principal component 2

(C) Italian cases and controls after exclusion of ancestry outliers.



PC1: principal component 1, PC2: principal component 2

(D) Spanish cases and controls after exclusion of ancestry outliers.

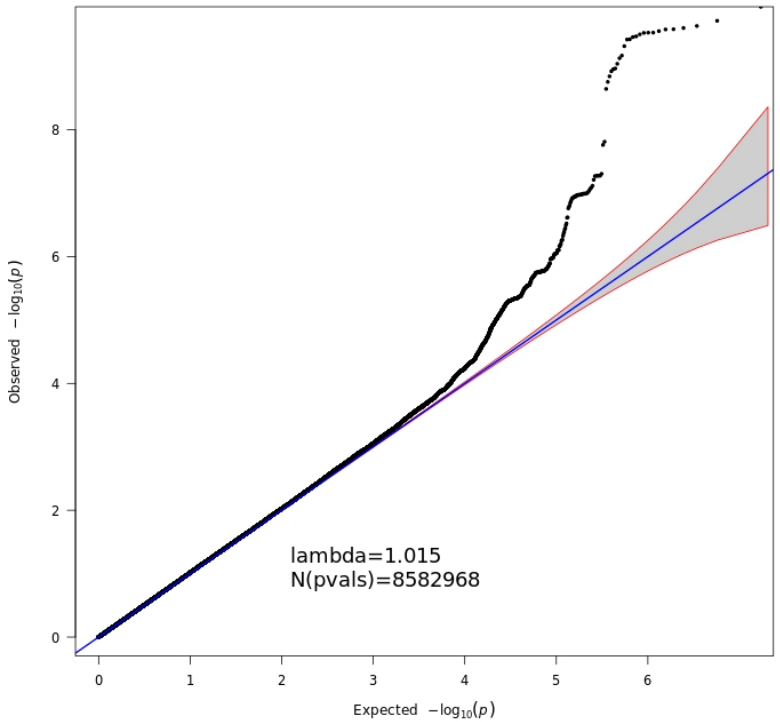


PC1: principal component 1, PC2: principal component 2

Figure S2. Quantile-quantile plots of meta-analyses of Italian and Spanish GWAS association summary statistics.

Only meta-analysed markers that passed the imputation score $R^2>0.6$ in both Italian and Spanish GWAS data sets were used for plotting. In quantile-quantile (Q-Q) plots the 2.5th and 97.5th centiles of the distribution under random sampling and the null hypothesis form the 95% concentration band. The genomic inflation factor lambda (λ**)** is defined as the ratio of the medians of the sample $\chi 2$ test statistics and the 1-d.f. $\chi 2$ distribution (0.455).[33] Two different analyses including covariates were conducted to control for potential population stratification (analysis I; see **Supplementary Methods**) as well as potential age and gender effects (analysis II; see **Supplementary Methods**).

A) Meta-analysis I (analysis definition see text above)



(B) Meta-analysis II (analysis definition see text above)

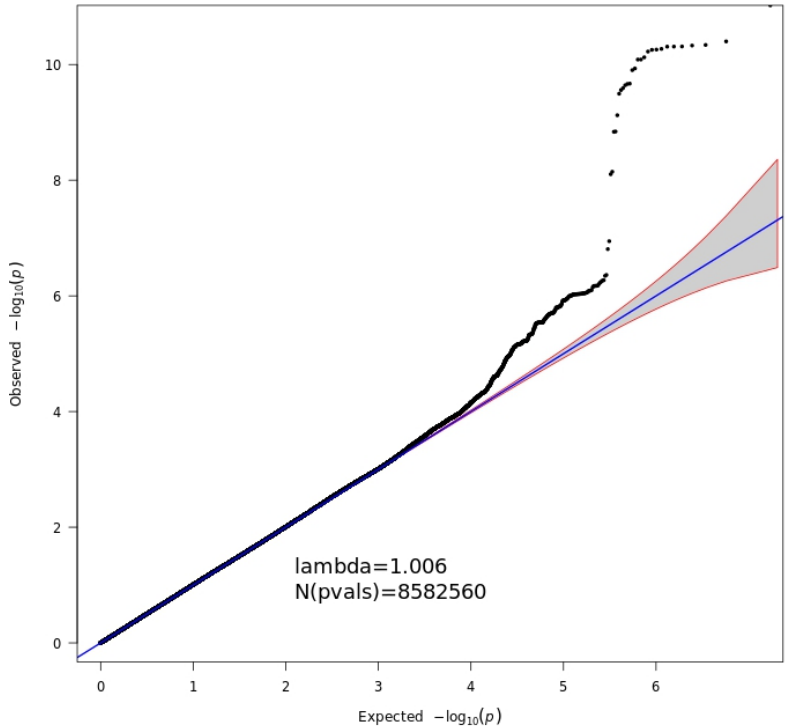


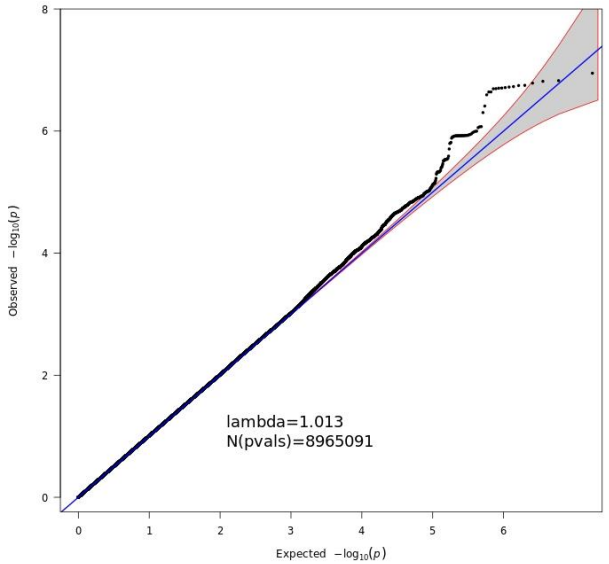
Figure S3. Quantile-quantile and Manhattan plots of Italian and Spanish GWAS association summary statistics.

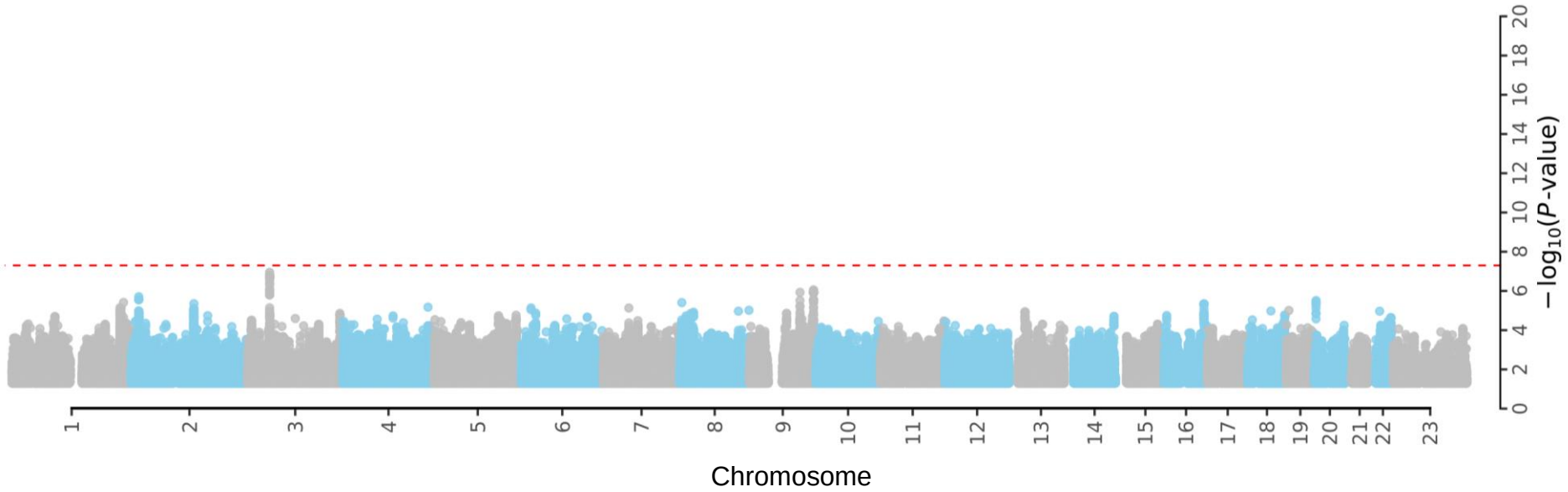
Only markers that passed the imputation score $R^2>0.6$ were used for plotting. In quantile-quantile (Q-Q) plots the 2.5th and 97.5th centiles of the distribution under random sampling and the null hypothesis form the 95% concentration band. The genomic inflation factor lambda (λ**)** is defined as the ratio of the medians of the sample $\chi 2$ test statistics and the 1-d.f. $\chi 2$ distribution (0.455).[33]

In the Manhattan plots of the association statistics from Italy and Spain, respectively, the red horizontal line indicates a genome-wide significance threshold of $P=5x10^{-8}$.

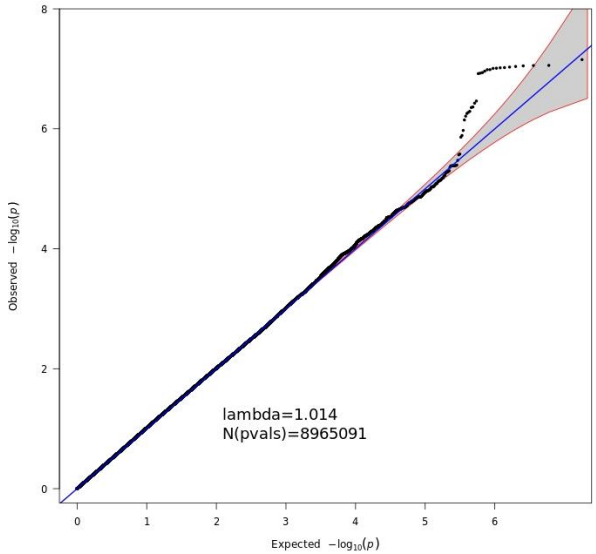
Two different analyses including covariates were conducted to control for potential population stratification (analysis I; see **Supplementary Methods**) as well as potential age and gender effects (analysis II; see **Supplementary Methods**).

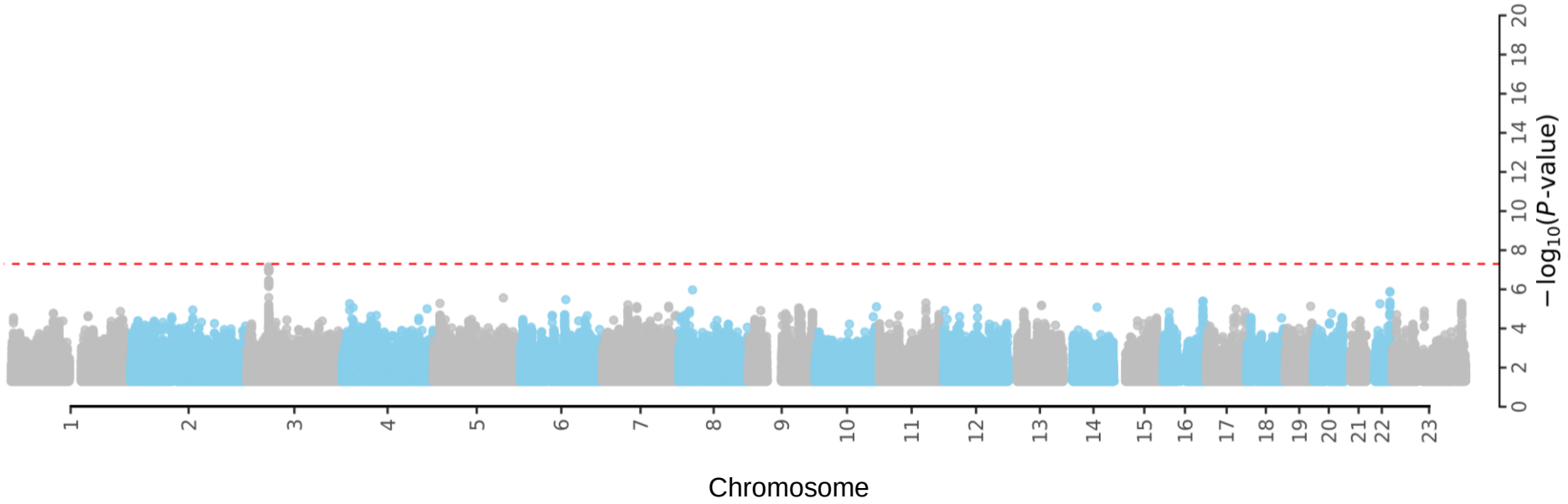
(A) Italy analysis I (analysis definition see text above).



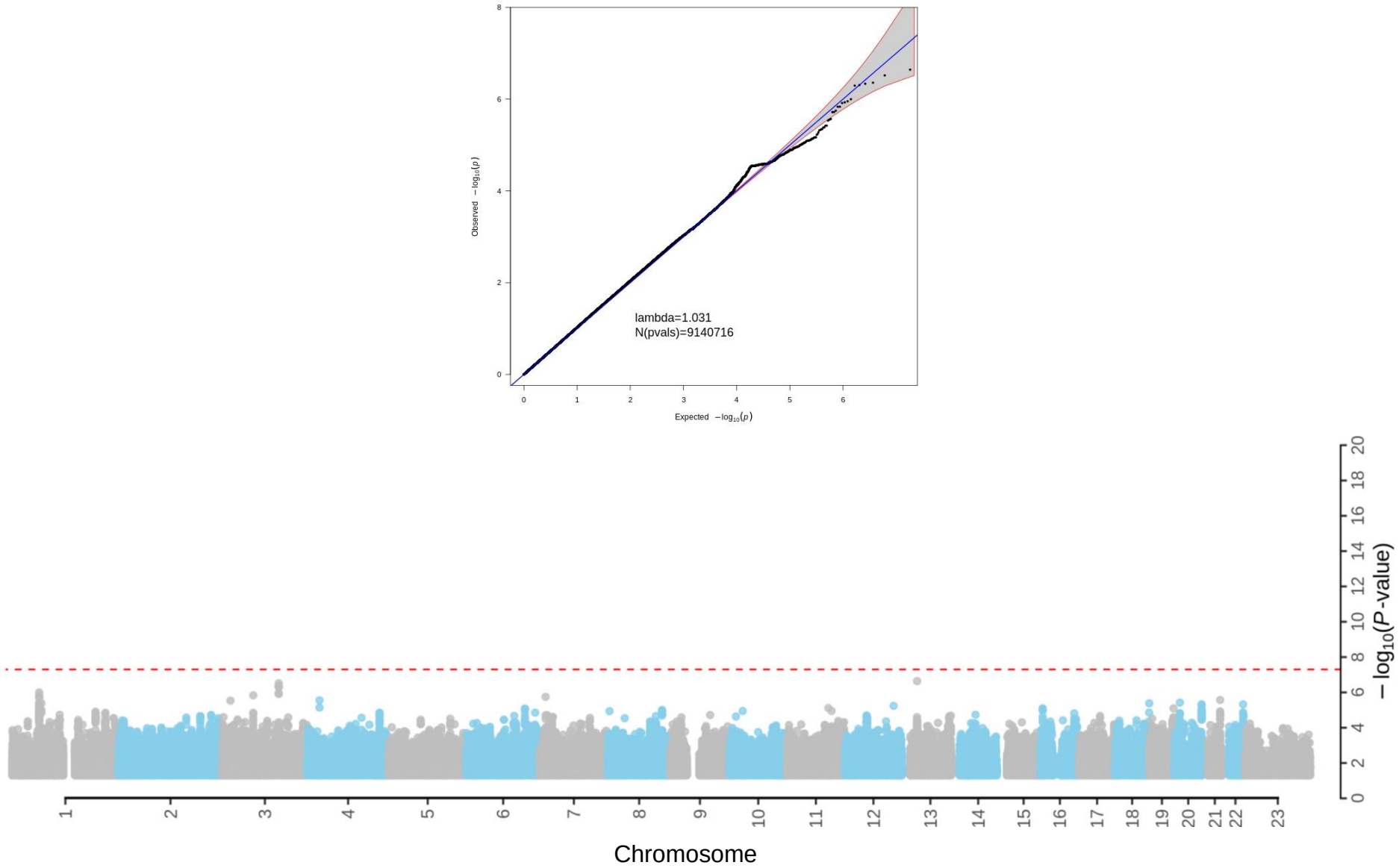


(B) Italy analysis II (analysis definition see text above).

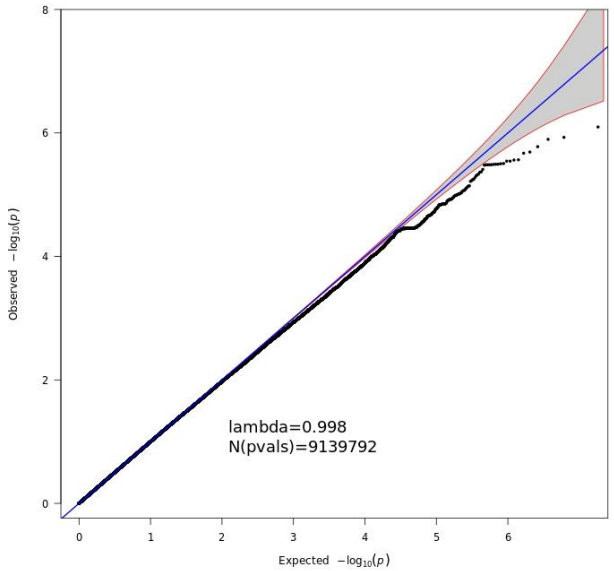




(C) Spain analysis I (analysis definition see text above).



(D) Spain analysis II (analysis definition see text above).



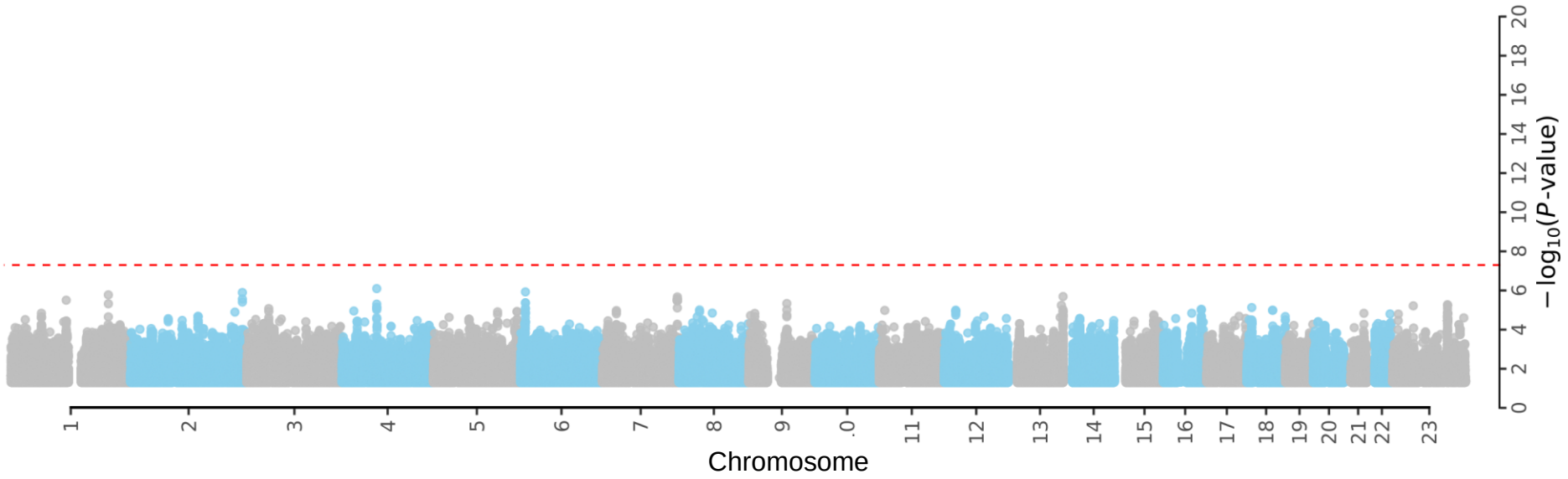


Figure S4. Genotype intensity cluster plots of genotyped SNP markers at 3p21.31 and 9q34.2. Genotype intensity cluster plots of 8 genotyped (i.e. non-imputed; MAF ≥1%) and most strongly associated SNP markers at 3p21.31 (5 SNPs selected per country, resulting in 8 unique SNPs) and the meta-analysis lead SNP rs657152 at 9q34.2 for 835 Italian cases and 1,255 Italian controls and 775 Spanish cases and 950 Spanish controls (see **Table S1, C**). Genotypes depicted in black were automatically set to missing by the Illumina Genomestudio software during initial genotype calling (prior to GWAS quality control). All intensity cluster plots have well separated and tightly bunched clusters for homozygous (red and blue) and heterozygous (green) genotypes. Original SNP identifiers from Genomestudio software are depicted.

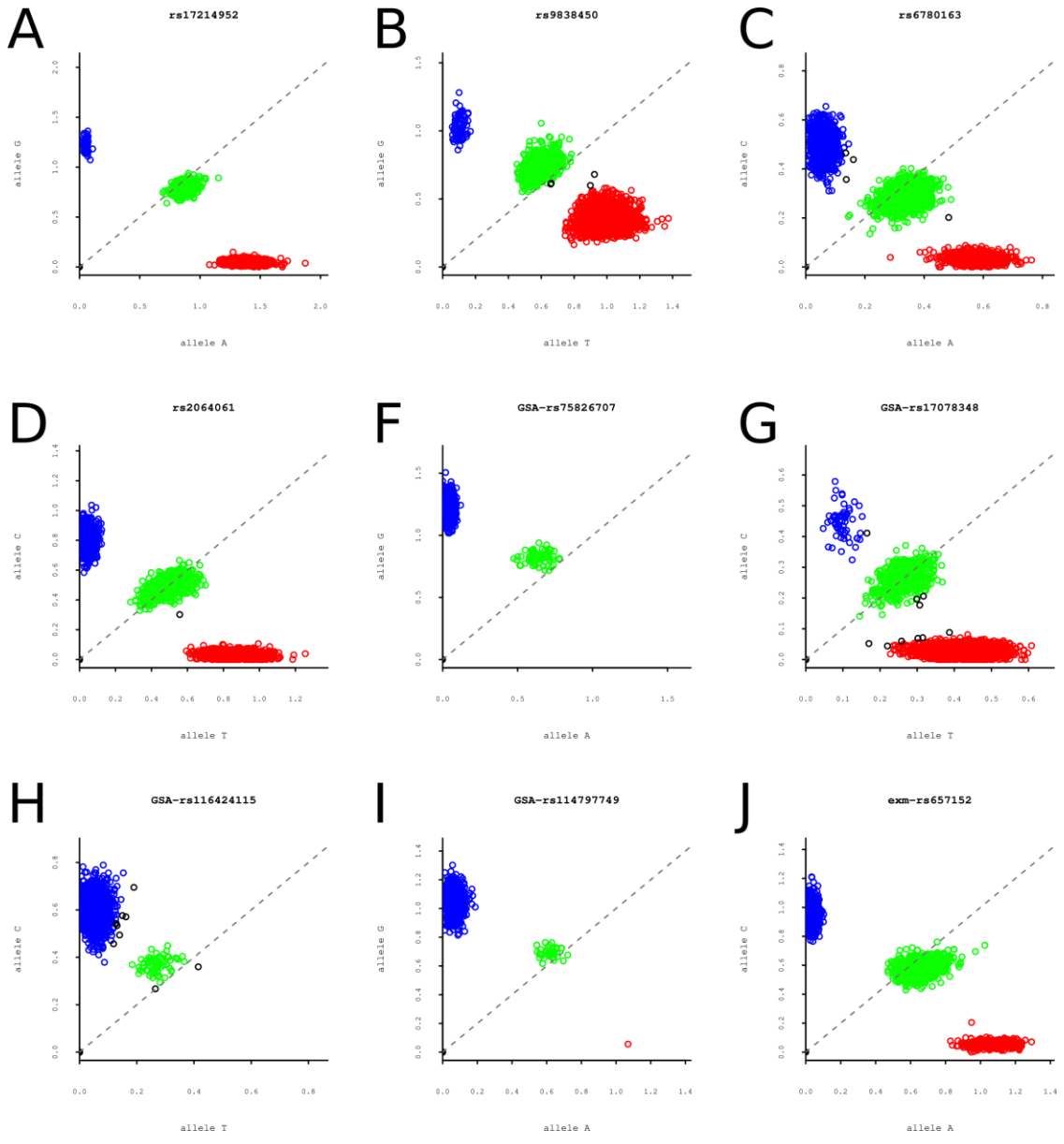


Figure S5. Manhattan plot of meta-analysis of Italian and Spanish GWAS association summary statistics.

Manhattan plot of the association statistics from the meta-analysis controlled for potential population stratification as well as potential age and gender effects (analysis II; see **Supplementary Methods**). The red horizontal line indicates a genome-wide significance threshold of $P=5x10^{-8}$.

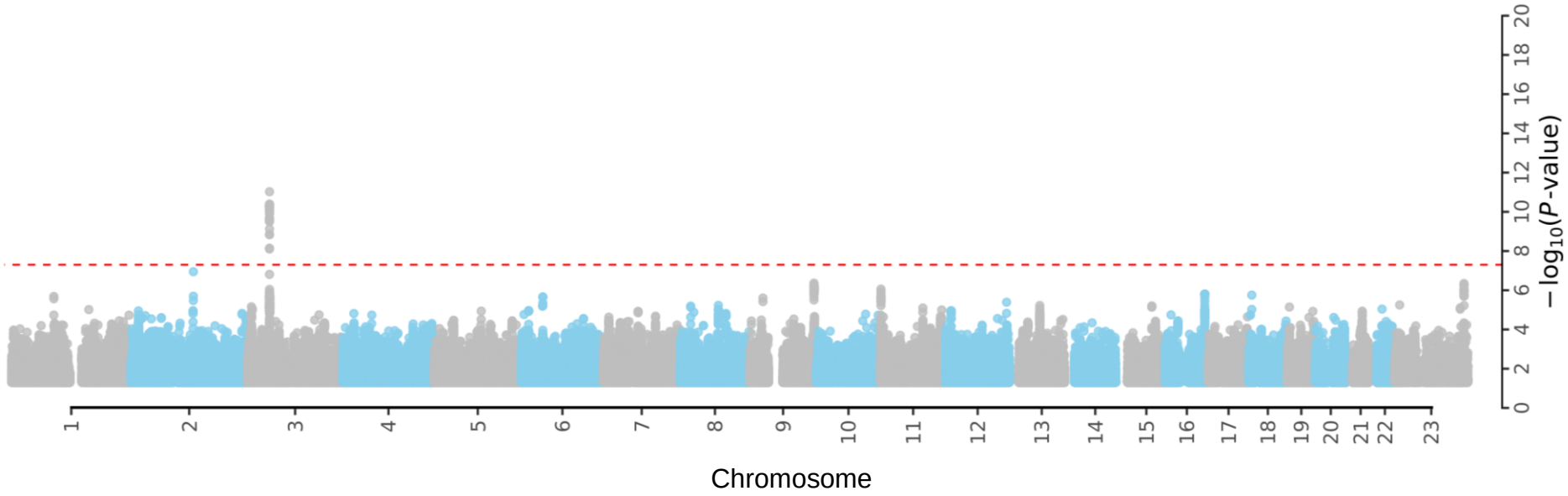


Figure S6. Manhattan plot of loci of suggestive evidence.

Manhattan plot of loci of suggestive evidence (blue horizontal line; $P<1\times10^{-5}$) for association with severe Covid-19 with respiratory failure in meta-analysis (analysis I, for analysis definition see **Supplementary Methods**) of Italian and Spanish GWAS association summary statistics. The red horizontal line indicates a genome-wide significance threshold of $P=5\times10^{-8}$.

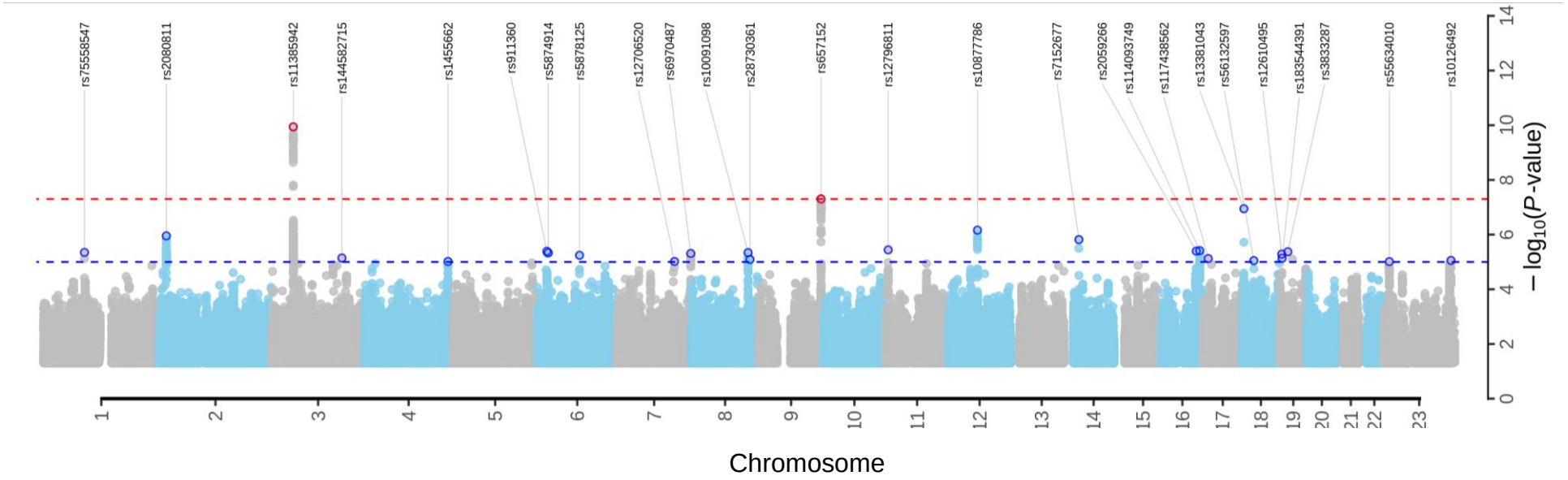


Figure S7. Expression levels of candidate genes of genome-wide significant loci in different tissues, blood cell types and lung single cell data.

Chromosome regions 3p21.31 and 9q34.2 (associated boundaries of **Table 2** and flanking sites) span several protein-coding genes, including *ABO, CCR1, CCR2, CCR3, CCR9, CXCR6, FYCO1, LARS2, LIMD1, LZTFL1, SACM1L, SLC6A20* and *XCR1*. In order to evaluate their potential role in SARS-Cov-2 pathogenesis, their mRNA gene expression data from publicly available sources were analyzed (see **Supplementary Methods**). (**A)** Figure shows consensus bulk mRNA gene expression results from Human Protein Atlas. Visualized expression values were gene-wise centered and z-score normalized; (**B-C)** Figures represent mean expression (visualized by color) of candidate genes and fraction of cells expressing those genes (visualized by the size of the dot). Lung and upper airway sc-RNA-Seq data from Vieira Braga et al[24] (B) contain mRNA expression levels in healthy nasal, bronchi, alveoli and parenchyma tissues. Data set (C) contains healthy human lung parenchyma sc-RNA-Seq results from Madissoon et al.[23] Processed sc-RNA-Seq datasets from both studies were retrieved from COVID-19 Cell Atlas.[22]

Consensus mRNA tissue gene expression data showed strong variation in candidate gene expression along multiple tissues and cell types. For example, in bulk mRNA-Seq data (A), *CXCR6* shows high expression in T cells when compared to other tissues and cell types, while *SLC6A20* shows higher expression in neutrophils. These findings are consistent with sc-RNA-Seq data of healthy lung cells (B), where *CXCR6* is expressed in lung residing T and NK cells and *SLC6A20* is expressed in nasal neutrophils. Expression data of *CCR1* gene is also consistent across datasets (B) and (C), this gene being highly expressed in myeloid cells, especially in lung residing macrophages. Notably, *LZTFL1* and *ABO* genes are expressed in alveolar and ciliated cells (of the respiratory epithelium in the lung) (B-C). Some of the candidate genes, including *CCR3* and *CCR9*, were not found to be noticeably expressed in these datasets.

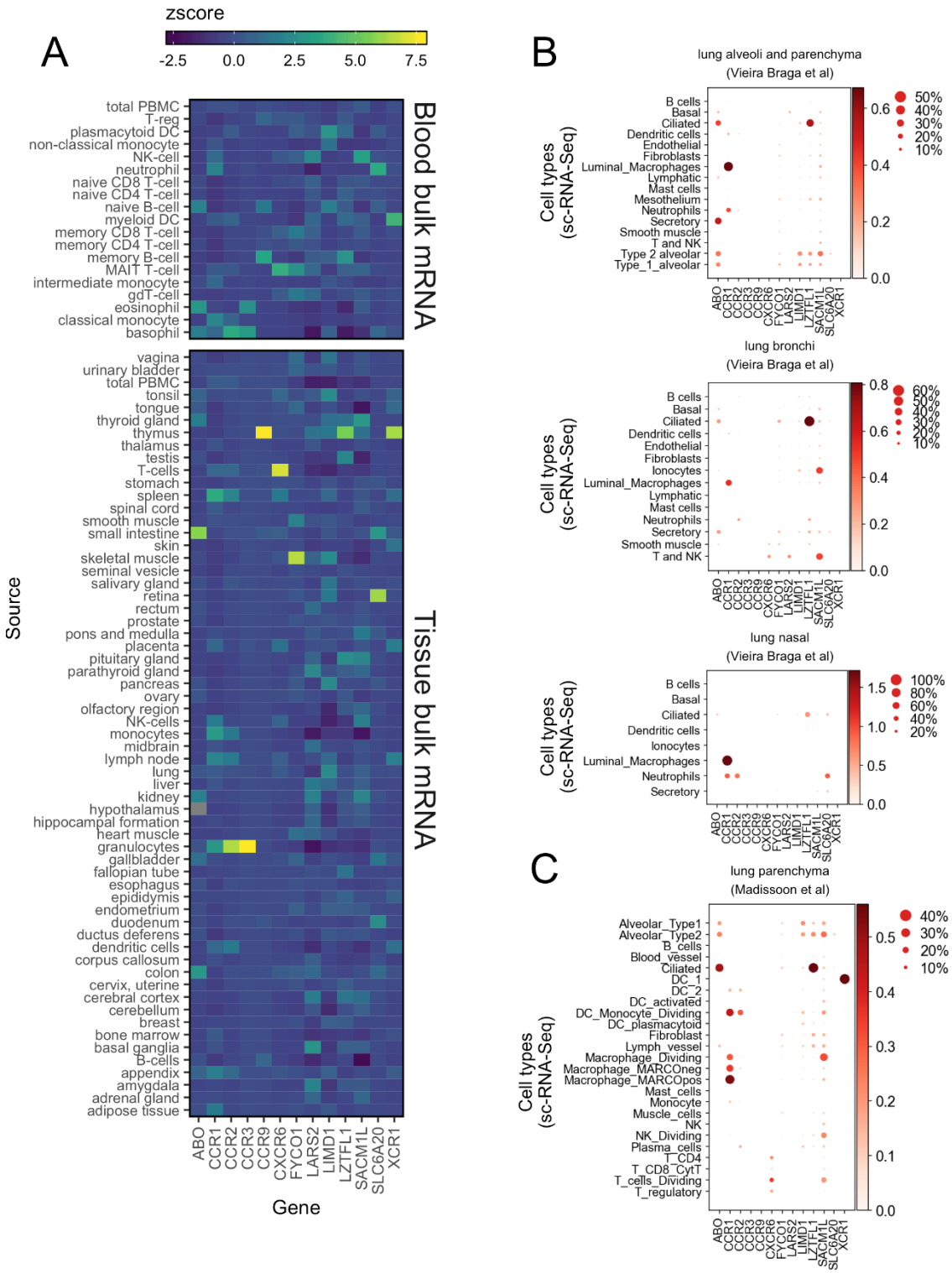
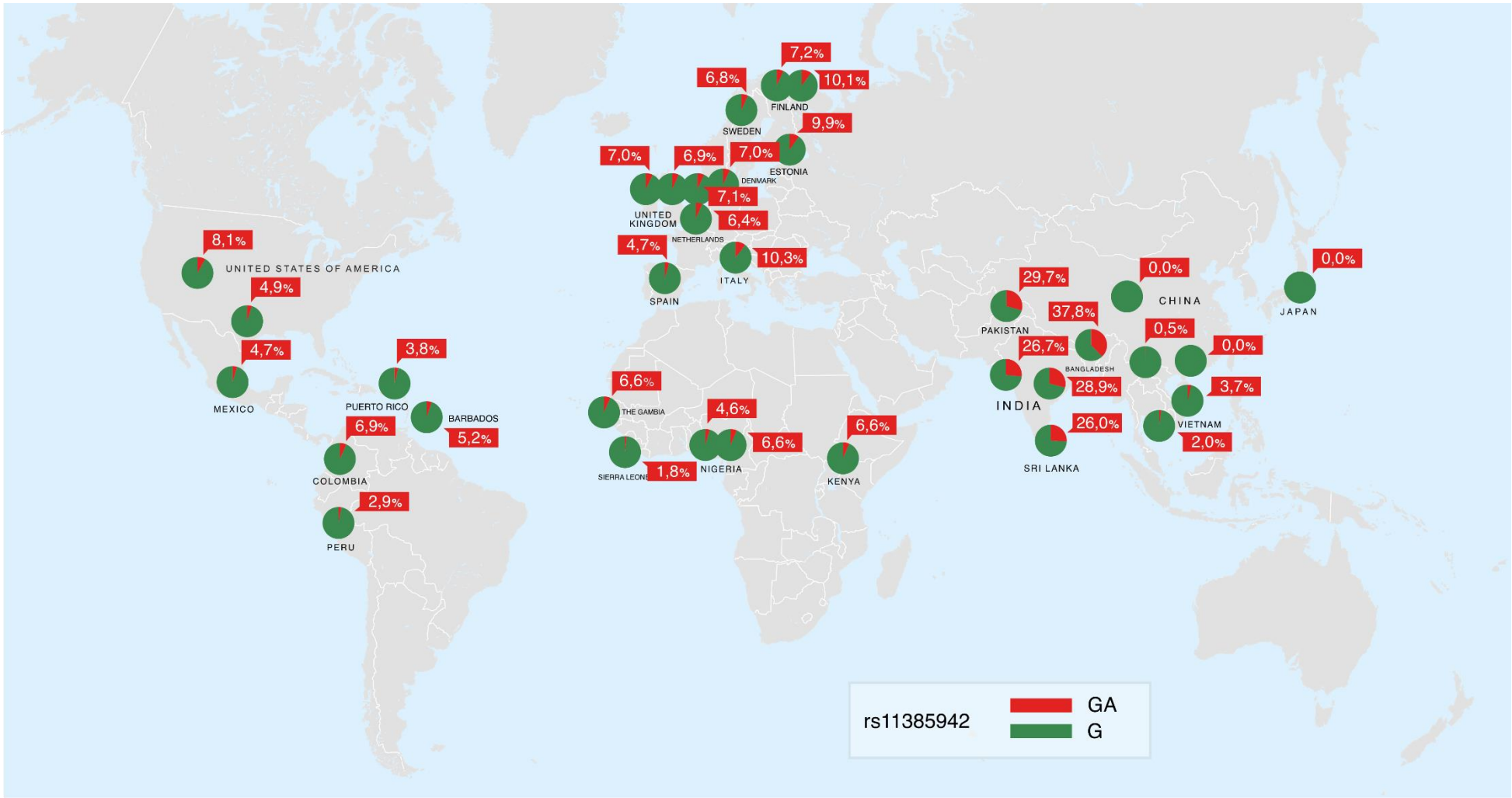

A
zscore
-2.5
0.0
2.5
5.0
7.5
Blood bulk mRNA
Tissue bulk mRNA
Source
Gene
ABO
CCR1
CCR2
CCR3
CCR9
CXCR6
FYCO1
LARS2
LIMD1
LZTFL1
SACM1L
SLC6A20
XCR1
B
lung alveoli and parenchyma
(Vieira Braga et al)
lung bronchi
(Vieira Braga et al)
lung nasal
(Vieira Braga et al)
Cell types
(sc-RNA-Seq)
C
lung parenchyma
(Madissoon et al)

Figure S8. World map of risk allele frequencies for rs11385942 and rs657152.

Based on populations of the 1,000 Genomes Project[1] (see also **Figure S1**), the Genome Aggregation Database[34] and the NCBI dbSNP.[35] The risk allele is colored in red, the protective allele is colored in green. rs11385942 is annotated as chr3:45834968-45834969:AAA:AA in dbSNPv153 and as chr3:45834967:GA:G in TOPMed imputation reference panel. Allele frequencies depicted for "SPAIN" and "ITALY" represent 1,000 Genomes Project frequencies for populations "Iberian Population in Spain" and "Toscani in Italia".


8,1%
UNITED STATES OF AMERICA
4,9%
4,7%
MEXICO
3,8%
PUERTO RICO
BARBADOS
5,2%
6,9%
COLOMBIA
2,9%
PERU
7,2%
6,8%
10,1%
FINLAND
SWEDEN
9,9%
ESTONIA
7,0%
6,9%
7,0%
DENMARK
7,1%
UNITED KINGDOM
6,4%
NETHERLANDS
4,7%
10,3%
ITALY
SPAIN
6,6%
THE GAMBIA
4,6%
6,6%
NIGERIA
SIERRA LEONE
1,8%
6,6%
KENYA
29,7%
0,0%
CHINA
0,0%
JAPAN
37,8%
PAKISTAN
0,5%
26,7%
BANGLADESH
0,0%
28,9%
INDIA
3,7%
26,0%
VIETNAM
2,0%
SRI LANKA
rs11385942
GA
G

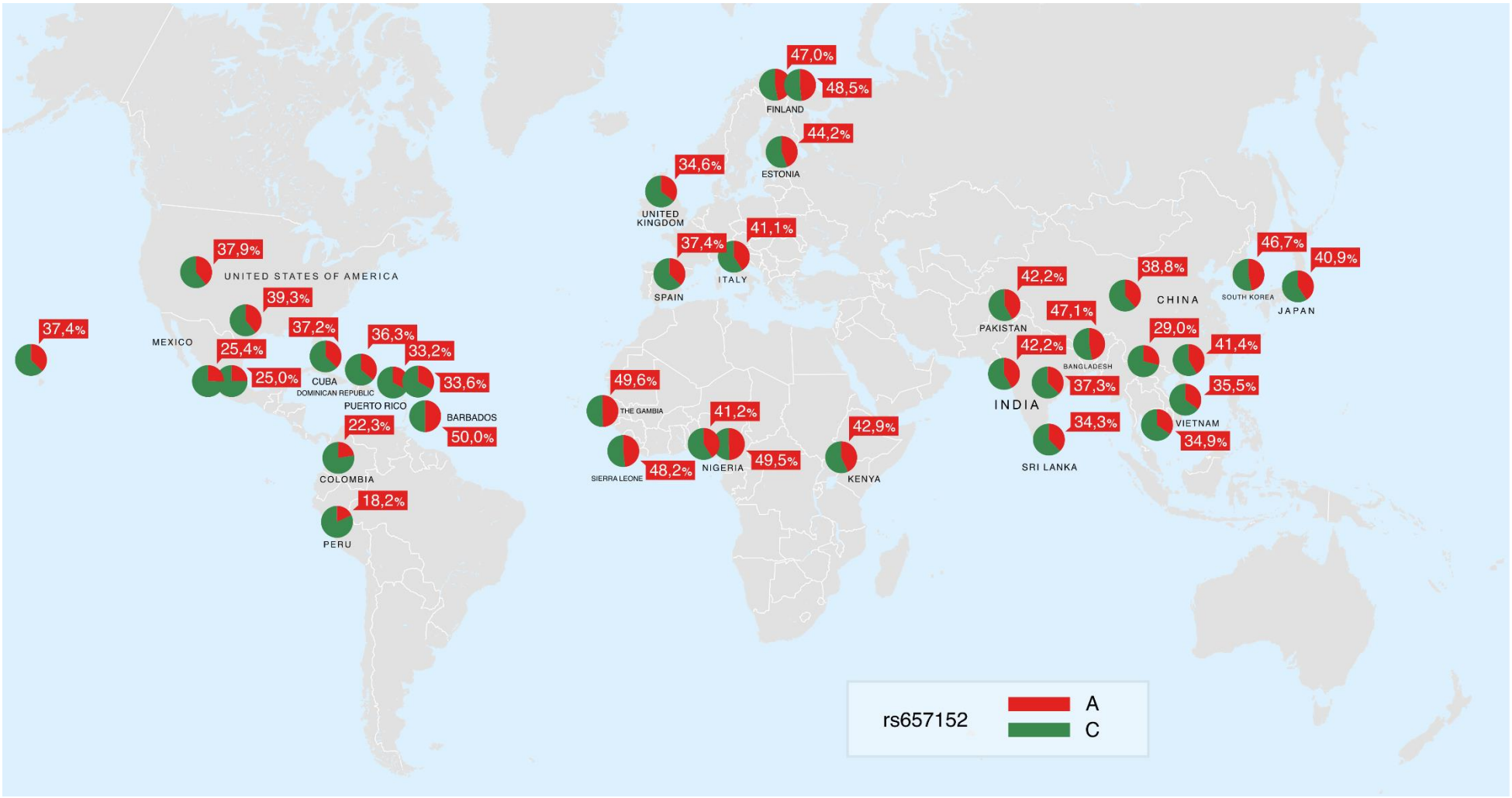
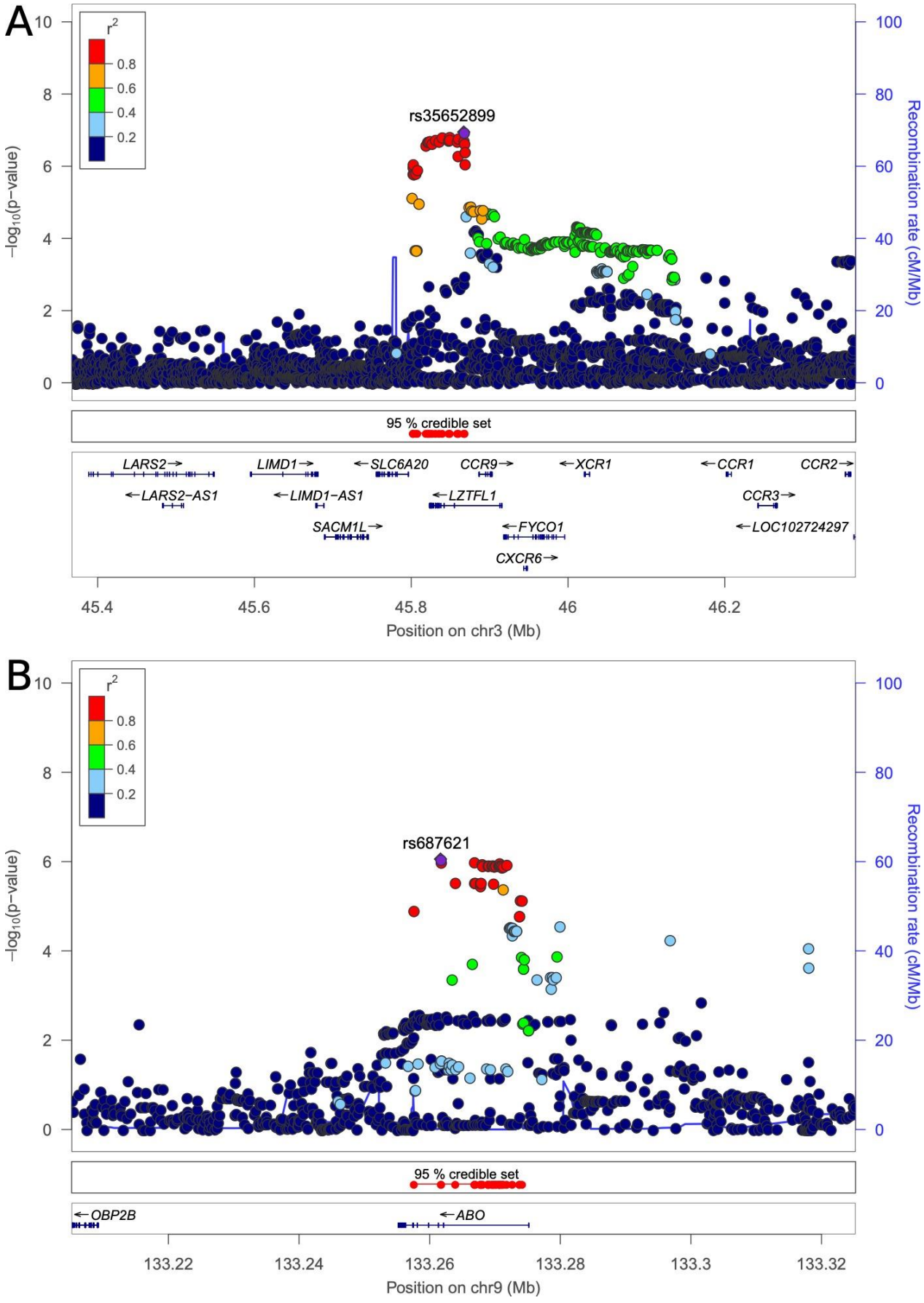

47,0%
48,5%
FINLAND
44,2%
ESTONIA
34,6%
UNITED KINGDOM
41,1%
37,4%
ITALY
SPAIN
37,9%
UNITED STATES OF AMERICA
39,3%
37,4%
MEXICO
37,2%
36,3%
33,2%
25,4%
25,0%
CUBA
DOMINICAN REPUBLIC
33,6%
PUERTO RICO
BARBADOS
50,0%
22,3%
COLOMBIA
18,2%
PERU
49,6%
THE GAMBIA
41,2%
SIERRA LEONE
48,2%
NIGERIA
49,5%
42,9%
KENYA
42,2%
PAKISTAN
47,1%
42,2%
BANGLADESH
INDIA
37,3%
34,3%
SRI LANKA
38,8%
CHINA
29,0%
41,4%
35,5%
VIETNAM
34,9%
46,7%
40,9%
SOUTH KOREA
JAPAN
rs657152
A
C

Figure S9. Regional association plots for loci 3p21.31 and 9q34.2 of Italian and Spanish GWAS association summary statistics.

Regional association plots for loci 3p21.31 and 9q34.2 of Italian and Spanish GWAS association summary statistics from analysis I (for analysis definition see **Supplementary Methods**). The purple diamond represents the most strongly associated SNP with severe Covid-19 and respiratory failure. The color illustrates linkage disequilibrium (LD) with the most strongly associated SNP, as shown in the color legend.

(A) Italy analysis I (for analysis definition see **Supplementary Methods**)



(B) Spain analysis I (for analysis definition see **Supplementary Methods**)

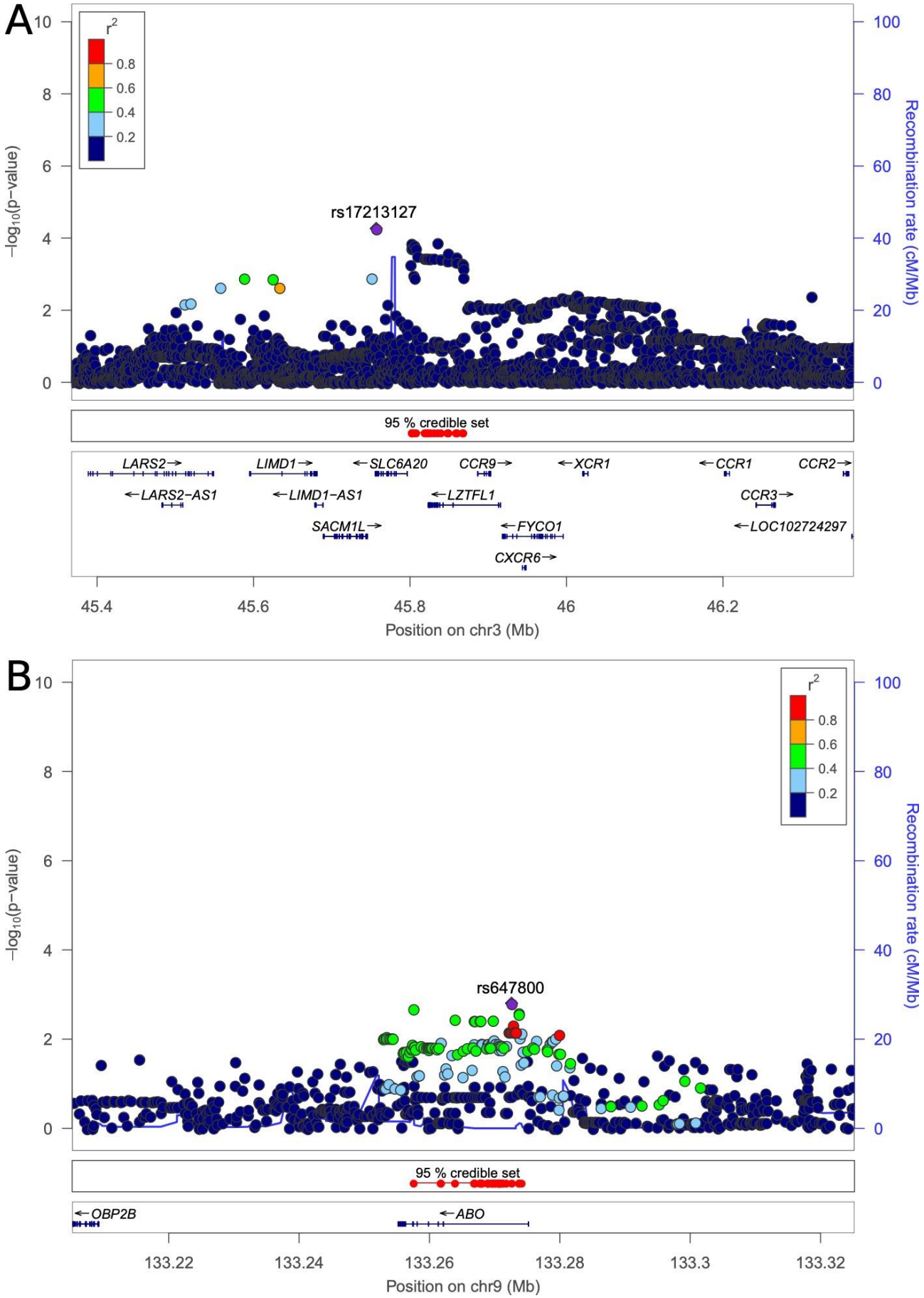


Figure S10. Regional association plot of the extended HLA region.

Regional association plot of the extended HLA region (chr6:25-35Mb) for (A) analysis (I) of GSA-based imputed HLA alleles across Italy and Spain and for (B) the Italian case-control cohort where NGS-based classical HLA allele information was available for a subset of the cohort (**Supplementary Methods**). SNPs are shown in light-blue circles and classical HLA alleles (at both 1st and 2nd field resolution) are shown as dark-blue triangles. There were no SNP or allele association signals meeting either the genome-wide (red dashed line) or the suggestive association significance threshold of $P=1x10^{-5}$ (orange dashed line).

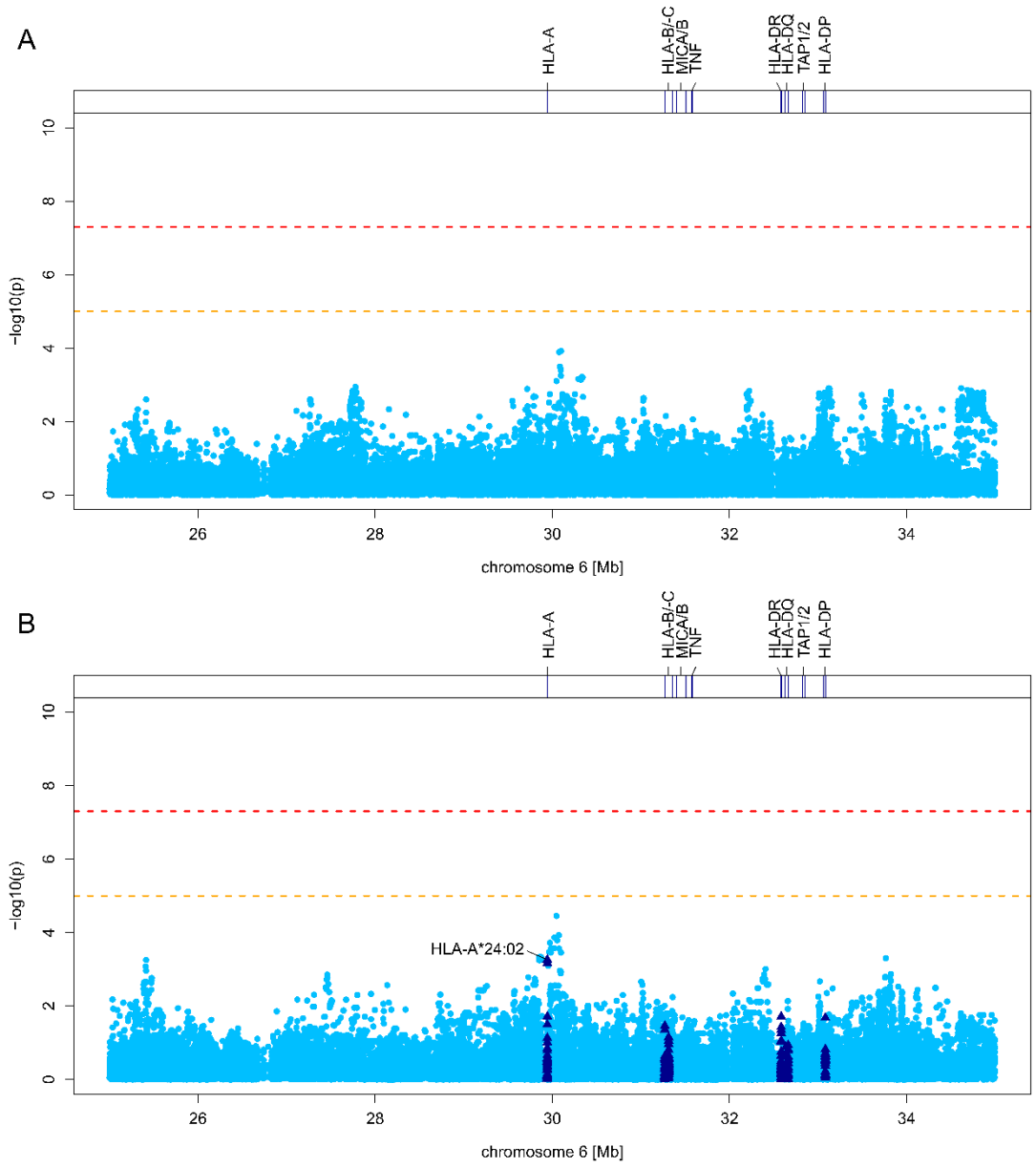
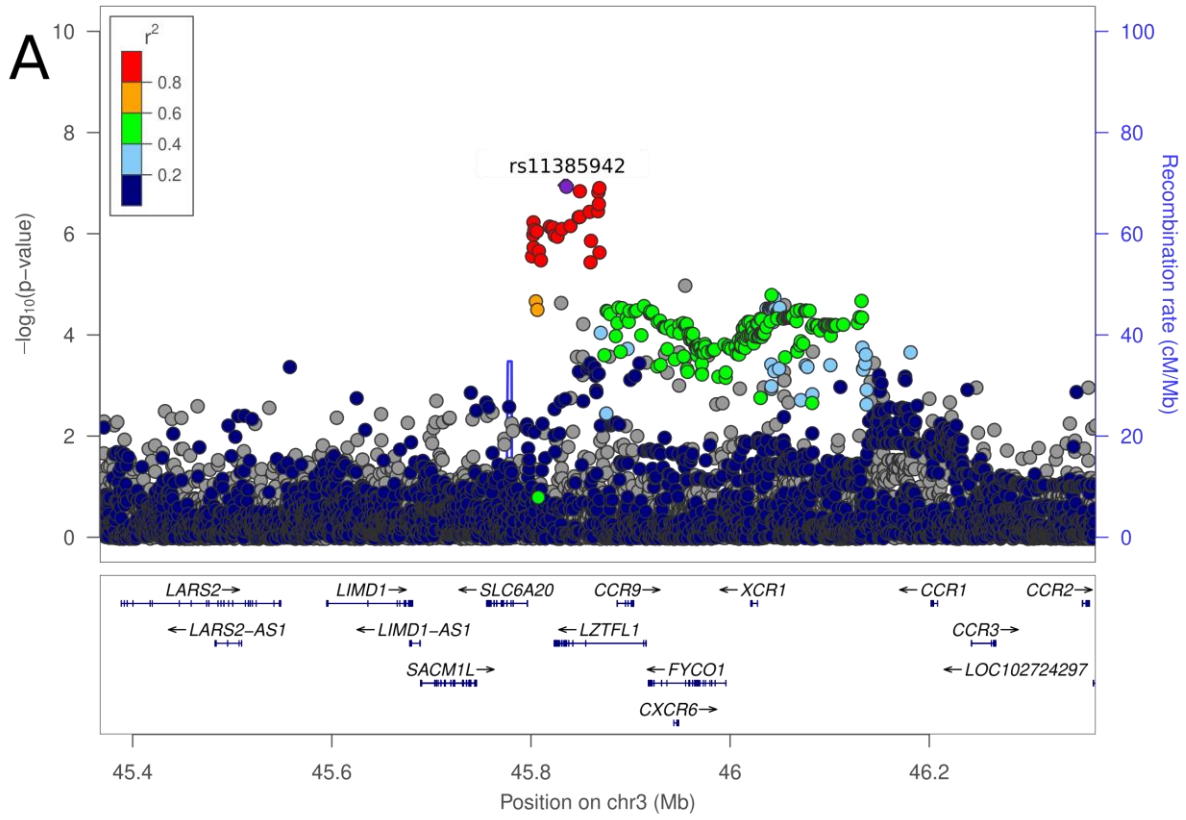
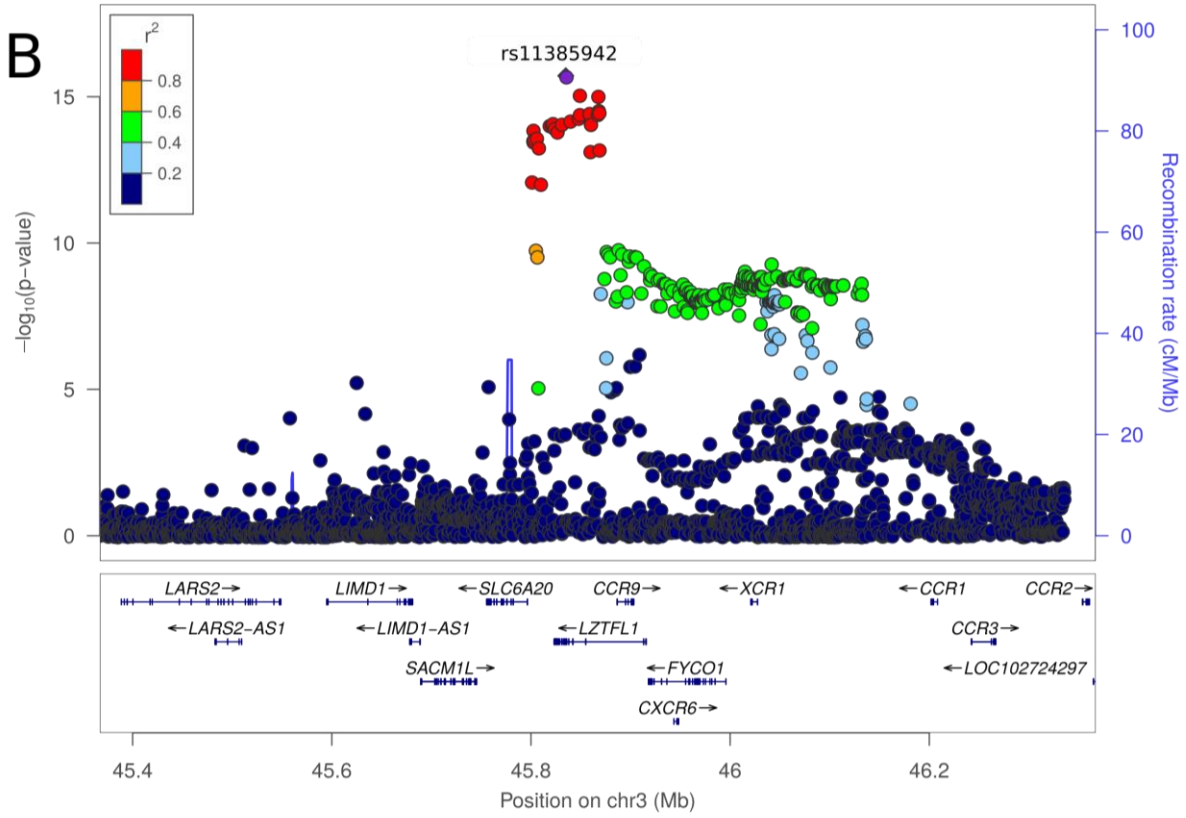


Figure S11. Regional association plots for locus 3p21.31 of the COVID-19 Host Genetics Initiative GWAS association summary statistics, along with a meta-analysis with our results. The purple diamond represents the most associated SNP at 3p21.31. The color illustrates linkage disequilibrium (LD) with the most associated SNP, as shown in the color legend. The summary statistics of the publicly available COVID-19 HGI (for details see **Supplementary Note**) GWAS meta-analyses round 2 (ANA5, susceptibility [affected vs. population]) was downloaded from https://www.covid19hg.org/results/ (analysis named "20200508-results-ANA5_ALL_inv_var_meta"; file named "COVID19_HGI_ANA5_20200513.txt.gz"; meta-analysis release date of May 15 2020) and plotted for 3p21.31 (Panel A). Further, a meta-analysis was performed using METAL (see **Supplementary Methods**) across the COVID19-HGI GWAS meta-analysis summary statistics and association results of our analysis I and results plotted similarly (Panel B). Finally, another meta-analysis was performed across the COVID19-HGI GWAS meta-analysis summary statistics and association results of our analysis II and plotted similarly (Panel C).

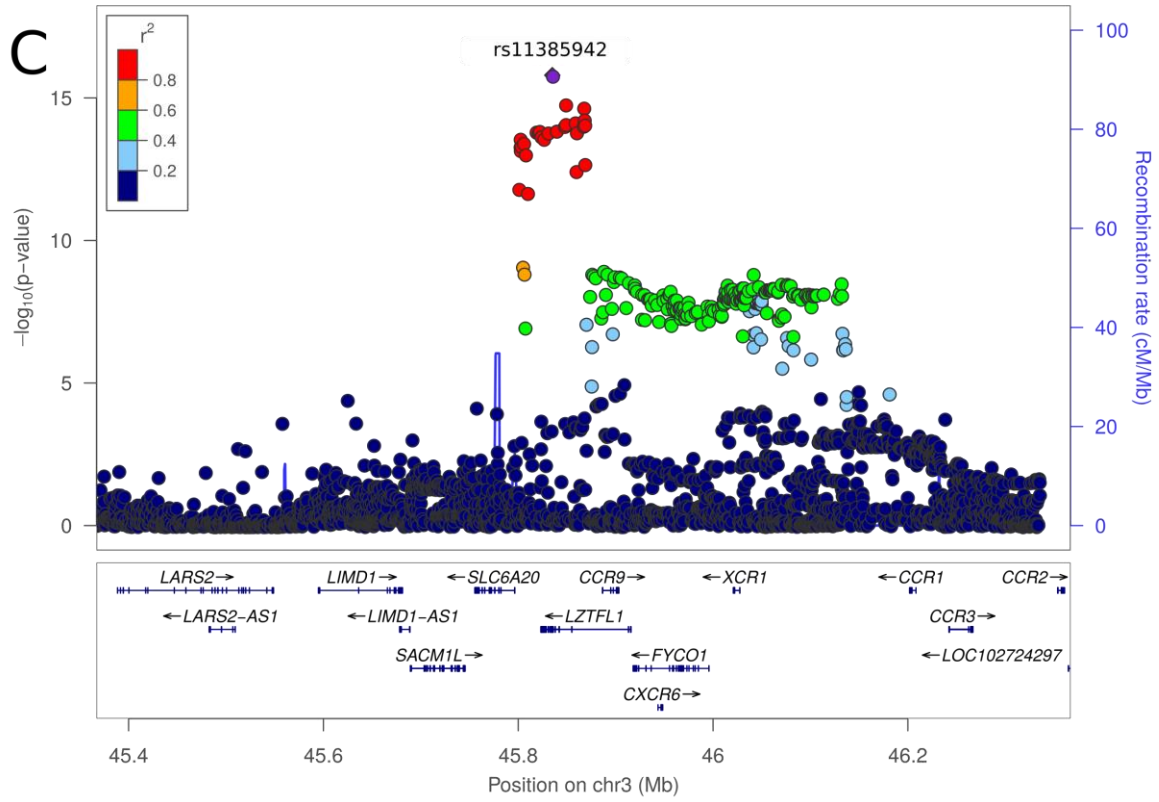
(A) Plotted COVID-19 HGI GWAS meta-analysis round 2 (ANA5) association results for locus 3p21.31. Most strongly associated SNP is rs11385942 ($P=1.10x10^{-7}$).



(B) Meta-analysis across the COVID-19 HGI GWAS meta-analysis summary statistics and association results of our analysis I for locus 3p21.31.



(C) Meta-analysis across the COVID-19 HGI GWAS meta-analysis summary statistics and association results of our analysis II for locus 3p21.31.



Supplementary Tables

Table S1. Patient and control GWAS panels before and after quality control.

(A) Number of samples fulfilling the inclusion criteria from each contributing center before GWAS quality control. (B) Number of cases and controls excluded from the final analysis in different quality control steps. For details, see **Supplementary Methods** section on genotype calling and SNP and sample quality control. (C) Total number of GWAS QCed cases and controls fulfilling the inclusion criteria.

(A) Total number of patients fulfilling inclusion criteria from each center before quality control.

	Eligible N
Fondazione IRCCS Cá Granda Ospedale Maggiore Policlinico, Milan	597
Humanitas Clinical and Research Center, IRCCS, Milano	154
UNIMIB School of Medicine, San Gerardo Hospital, Monza	200
Hospital Clínic and IDIBAPS, Barcelona	56
Hospital Universitario Vall d'Hebron, Barcelona	337
Hospital Universitario Ramón y Cajal, Madrid	298
Donostia University Hospital, San Sebastian	338

(B) Overview of genotyped cases and controls from each country with reasons for exclusion.

	Italy		**Spain**	
	Cases	**Controls**	**Cases**	**Controls**
Pre-QC	951	1,394	1,029	987
Pre-QC totals (cases+controls)	2,345		2,016	
Missingness outliers	13		10	
Heterozygosity outliers	20		15	
PCA outliers	152		201	
Duplicates	19		69	
Relatives	57		27	
Total unique removed*	255		291	
Post-QC totals	2,090		1,725	
	835	1,255	775	950

*The total number of unique samples removed from analysis is smaller than the sum of reasons for exclusion since some samples may have several.
QC: quality control.

(C) Total number of patients fulfilling inclusion criteria after quality control.

Table shows the number, gender and age of patients and controls included in the study, after quality control, sorted according to the two geographic panels of Italy and Spain.

	Italy		Spain	
	Cases (n=835)	**Controls (n=1,255)**	**Cases (n=775)**	**Controls (n=950)**
Gender female, n (%)	249 (30)	495 (39)	265 (34)	310 (33)*
Age, median (IQR)	65 (56-75)	49 (33-59)	67 (58-75)	44 (33-50)**

*Missing gender data of two controls
**Missing age data of 25 controls

Table S2: Frequency data for lead SNPs in clinical subgroups.

Sample sizes are provided in brackets.

			Frequency data from Table 2				**Coronary artery disease Risk allele frequencies**				**Diabetes Risk allele frequencies**				**Hypertension Risk allele frequencies**			
			Italy all		**Spain all**		**Italy**		**Spain***		**Italy**		**Spain***		**Italy**		**Spain***	
		Risk allele	**Cases (835)**	**Controls (1255)**	**Cases (775)**	**Controls (950)**	**CAD (71)**	**No CAD (764)**	**CAD (68)**	**No CAD (705)**	**DM (125)**	**No DM (710)**	**DM (182)**	**No DM (591)**	**HT (346)**	**No HT (489)**	**HT (365)**	**No HT (408)**
3p21.31	rs11385942	GA	0.14	0.09	0.09	0.05	0.09	0.15	0.12	0.09	0.14	0.14	0.08	0.10	0.13	0.15	0.09	0.10
9q34.2	rs657152	A	0.42	0.35	0.42	0.35	0.47	0.42	0.48	0.41	0.42	0.42	0.46	0.41	0.43	0.42	0.42	0.42

CAD: Coronary artery disease; DM: diabetes mellitus; HT: hypertension; *Information missing on n=2.

Table S3. Risk allele frequencies for rs11385942 and rs657152 for six external and study independent GWAS control data sets.

Risk allele frequencies for rs11385942 and rs657152 for six external and study independent TOPMed-imputed Spanish and Italian population-based GWAS control data sets from different genotyping platforms.

SEE EXCEL FILE

Table S4. Test for association assuming a recessive and a heterozygous model in the meta-analyses.

SEE EXCEL FILE

Table S5. Genomic loci showing suggestive evidence with severe Covid-19 and respiratory failure in the meta-analysis.

Genomic loci showing suggestive evidence ($P<1x10^{-5}$) with severe Covid-19 and respiratory failure in meta-analysis I. All listed data represent output from the FUMA analysis, see **Supplementary Methods** section for further details.

SEE EXCEL FILE

Table S6. Variants in 95% fine-mapped credible sets at 3p21.31 and 9q34.2.

SEE EXCEL FILE

Table S7. Tissue specific eQTL data of lead SNPs and proxy variants.

SEE EXCEL FILE

Table S8. Severity analysis for rs11385942 and rs657152 at chromosome 3p21.31 and 9q34.2.

SEE EXCEL FILE

Table S9. ABO blood group analysis results.

SEE EXCEL FILE

Table S10. HLA allele association analysis.

SEE EXCEL FILE

References

1. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. Nature 2015;526:68-74.
2. Dubois PC, Trynka G, Franke L, et al. Multiple common variants for celiac disease influencing immune gene expression. Nat Genet 2010;42:295-302.
3. Bentham J, Morris DL, Graham DSC, et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nat Genet 2015;47:1457-64.
4. Myocardial Infarction Genetics C, Kathiresan S, Voight BF, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet 2009;41:334-41.
5. Julia A, Gonzalez I, Fernandez-Nebro A, et al. A genome-wide association study identifies SLC8A3 as a susceptibility locus for ACPA-positive rheumatoid arthritis. Rheumatology (Oxford) 2016;55:1106-11.
6. Lopez-Isac E, Acosta-Herrera M, Kerick M, et al. GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. Nat Commun 2019;10:4955.
7. Obon-Santacana M, Vilardell M, Carreras A, et al. GCAT|Genomes for life: a prospective cohort study of the genomes of Catalonia. BMJ Open 2018;8:e018324.
8. Galvan-Femenia I, Obon-Santacana M, Pineyro D, et al. Multitrait genome association analysis identifies new susceptibility genes for human anthropometric variation in the GCAT cohort. J Med Genet 2018;55:765-78.
9. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-75.
10. Price AL, Weale ME, Patterson N, et al. Long-range LD can confound genome scans in admixed populations. Am J Hum Genet 2008;83:132-5; author reply 5-9.
11. Abraham G, Qiu Y, Inouye M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. Bioinformatics 2017;33:2776-8.
12. Das S, Forer L, Schonherr S, et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48:1284-7.
13. Loh PR, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet 2015;47:284-90.
14. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190-1.
15. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun 2017;8:1826.
16. Benner C, Spencer CC, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics 2016;32:1493-501.
17. Bugert P, Rink G, Kemp K, Kluter H. Blood Group ABO Genotyping in Paternity Testing. Transfus Med Hemother 2012;39:182-6.
18. Aguet F, Barbeira AN, Bonazzola R, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. bioRxiv 2019:787903.
19. Jansen R, Hottenga JJ, Nivard MG, et al. Conditional eQTL analysis reveals allelic heterogeneity of gene expression. Hum Mol Genet 2017;26:1444-51.
20. Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. Science 2015;347:1260419.
21. Consortium F, the RP, Clst, et al. A promoter-level mammalian expression atlas. Nature 2014;507:462-70.
22. Sungnak W, Huang N, Becavin C, et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. Nat Med 2020;26:681-7.

23. Madissoon E, Wilbrey-Clark A, Miragaia RJ, et al. scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. Genome Biol 2019;21:1.
24. Vieira Braga FA, Kar G, Berg M, et al. A cellular census of human lungs identifies novel cell states in health and in asthma. Nat Med 2019;25:1153-63.
25. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 2018;19:15.
26. Pierini F, Lenz TL. Divergent Allele Advantage at Human MHC Genes: Signatures of Past and Ongoing Selection. Mol Biol Evol 2018;35:2145-58.
27. Arora J, Pierini F, McLaren PJ, Carrington M, Fellay J, Lenz TL. HLA Heterozygote Advantage against HIV-1 Is Driven by Quantitative and Qualitative Differences in HLA Allele-Specific Peptide Presentation. Mol Biol Evol 2020;37:639-50.
28. Chowell D, Krishna C, Pierini F, et al. Evolutionary divergence of HLA class I genotype impacts efficacy of cancer immunotherapy. Nat Med 2019;25:1715-20.
29. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res 2020.
30. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall 2017;1:33-46.
31. Zehender G, Lai A, Bergna A, et al. Genomic characterization and phylogenetic analysis of SARS-COV-2 in Italy. J Med Virol 2020.
32. Grifoni A, Weiskopf D, Ramirez SI, et al. Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. Cell 2020.
33. Devlin B, Roeder K. Genomic control for association studies. Biometrics 1999;55:997-1004.
34. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581:434-43.
35. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29:308-11.